

ECOLE DOCTORALE DES SCIENCES DE LA VIE,
SANTÉ, AGRONOMIE, ENVIRONNEMENT
N° d'ordre 559

Thèse

Présentée à l'Université Blaise Pascal pour l'obtention du grade de

DOCTEUR D'UNIVERSITE
(Spécialité : Génomique et Ecologie Microbienne)

Présentée et soutenue publiquement le
07 Novembre 2011

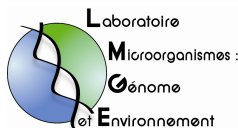
Eric DUGAT-BONY

CARACTERISATION DES CAPACITES METABOLIQUES DES POPULATIONS MICROBIENNES IMPLIQUEES DANS LES PROCESSUS DE BIOREMEDIATION DES CHLOROETHENES PAR DES APPROCHES MOLECULAIRES HAUT DEBIT : LES BIOPUCES ADN FONCTIONNELLES

Jury composé de :

- Rapporteurs : Wafa ACHOUAK (CR CNRS, LEMiRE, UMR 6191 CEA Cadarache, St Paul Lez Durance)
Philippe BERTIN (Pr., GMGM, UMR 7156, Université de Strasbourg)
- Examineurs : Richard BONNET (PU-PH, JE2526, Université d'Auvergne, Clermont-Ferrand)
Timothy VOGEL (Pr., Laboratoire Ampère, UMR 5005, Université Claude Bernard de Lyon)
Pierre PEYRET (Pr., LMGE, UMR 6023, Université d'Auvergne, Clermont-Ferrand)
- Directeur : Corinne BIDERRE-PETIT (CR CNRS, LMGE, UMR 6023, Université Blaise Pascal, Clermont-Ferrand)
- Invités : Jean-Yves RICHARD (Responsable R&D, SITA Remédiation, Suez Environnement, Meyzieu)
Cécile GRAND (Ingénieur ADEME, Angers)

Laboratoire « Microorganismes : Génome et Environnement »
Unité mixte de recherche CNRS 6023



Caractérisation des capacités métaboliques des populations microbiennes impliquées dans les processus de bioremédiation des chloroéthènes par des approches moléculaires haut débit : les biopuces ADN fonctionnelles

Résumé :

Les chloroéthènes sont les polluants majeurs des eaux souterraines et des nappes phréatiques. De par leur toxicité et leur effet cancérigène, ils représentent une préoccupation majeure pour les autorités publiques et sanitaires. La restauration des sites contaminés est possible par des techniques de dépollution biologique impliquant les microorganismes (bioremédiation microbienne). Cependant, la réussite des traitements dépend à la fois des conditions physico-chimiques du site pollué et des capacités de dégradation de la microflore indigène. Ainsi, pour optimiser les processus de décontamination, l'identification et le suivi des différentes populations microbiennes sont indispensables avant et pendant le traitement. Les biopuces ADN fonctionnelles (FGA, Functional Gene Array), outils moléculaires haut débit, sont particulièrement bien adaptées pour des applications en bioremédiation. Leur élaboration nécessite de disposer de logiciels performants pour le design de sondes qui combinent à la fois une forte sensibilité, une très bonne spécificité et un caractère exploratoire, ce dernier étant indispensable pour la détection des séquences connues mais surtout de celles encore jamais décrites au sein d'échantillons environnementaux. Un nouveau logiciel, autorisant la sélection de sondes combinant tous ces critères, a été développé et nommé HiSpOD. Son utilisation pour la construction d'une FGA dédiée aux voies de biodégradation des chloroéthènes a permis d'évaluer l'effet de traitements de biostimulation sur la microflore indigène pour plusieurs sites industriels contaminés. Les données révèlent différentes associations entre microorganismes déhalorespirants qui sont fonction des paramètres environnementaux.

Mots clés : bioremédiation, chloroéthènes, microorganismes, biopuce ADN.

Characterization of microbial populations' capacities involved in chloroethenes bioremediation processes using high-throughput molecular tools: functional DNA microarrays

Abstract:

Chlorinated solvents are among the most frequent contaminants found in groundwater and subsurface ecosystems. Because of their high toxicity and carcinogenicity, they represent a serious risk for human health and the environment. Thus, such polluted sites need a rehabilitation treatment. Among remediation solutions, microbial bioremediation represents a less invasive and expensive alternative than physico-chemical treatments. However, the process efficiency greatly depends on the environmental conditions and the microbial populations' biodegradation capacities. Therefore, bioremediation treatment optimization requires the identification and monitoring of such capacities before and during the treatment. Functional Gene Arrays (FGA), by profiling environmental communities in a flexible and easy-to-use manner, are well adapted for an application in bioremediation. But, constructing efficient microarrays dedicated to microbial ecology requires a probe design step allowing the selection of highly sensitive, specific and explorative oligonucleotides. After a detailed state of the art on probe design strategies suitable for microbial ecology studies, we present new software, called HiSpOD, generating efficient explorative probes for FGA dedicated to environmental applications. Finally, this bioinformatics tool was used to construct a FGA targeting most genes involved in chloroethenes biodegradation pathways which allowed the evaluation of biostimulation treatments conducted on indigenous bacterial populations for several industrial contaminated sites.

Keywords: bioremediation, chloroethenes, microorganisms, DNA microarrays.

Remerciements

Je souhaite exprimer mes remerciements à Christian Amblard, directeur du Laboratoire Microorganismes : Génome et Environnement (UMR CNRS 6023), pour m'avoir accueilli au sein de son laboratoire et permis de réaliser cette thèse dans des conditions idéales.

Mes plus sincères et chaleureux remerciements vont ensuite tout naturellement à Corinne Petit, mon encadrante durant ces trois années de thèse, pour son incroyable disponibilité et son efficacité qui sont pour beaucoup dans l'aboutissement de ce travail. Merci pour ta gentillesse, ton hospitalité, pour les agréables moments passés en compagnie de ta famille, merci de m'avoir tant donné... quel beau duo, n'est-ce pas ?

Un grand merci également à Pierre Peyret, responsable de l'équipe GIIM, pour la confiance qu'il m'a accordé depuis plus de cinq années maintenant et qui est pour beaucoup dans la passion de la recherche qui m'habite désormais et qui, je l'espère, continuera de me procurer ce sentiment encore longtemps. Merci pour votre soutien sans faille, vos précieux conseils, la grande liberté que vous m'avez laissé pour mener à bien mon projet, l'environnement de travail dont j'ai bénéficié durant toutes ces années et tout ce que vous avez fait pour moi.

Mes remerciements vont aussi à Eric Peyretailade, qui au travers de l'encadrement de mon Master mais également de son soutien et de l'intérêt qu'il a porté à tous les travaux auxquels j'ai pu participer, a également fortement contribué au bon déroulement de ma thèse. Merci également pour m'avoir transmis de précieux conseils quant à mon activité d'enseignement, pour ta bonne humeur quotidienne et pour tous les bons moments que nous avons partagé, notamment notre virée aux Karellis que je n'oublierai jamais.

Un grand merci également à Delphine Boucher, pour tous ses conseils, sa gentillesse, les agréables moments passés ensemble et pour la relecture finale de mon manuscrit. Quand tu veux pour une nouvelle choucroute...

Je souhaite exprimer mes remerciements les plus sincères à Wafa Achouak et Philippe Bertin pour m'avoir fait l'honneur d'accepter d'être rapporteurs de cette thèse ainsi qu'à Richard Bonnet, Timothy Vogel, Jean-Yves Richard et Cécile Grand pour avoir accepté d'examiner ce travail et de faire partie de mon jury. Merci également à l'Agence De l'Environnement et de la Maîtrise de l'Energie (ADEME) pour la bourse de thèse qu'elle m'a attribué.

Merci à tous les membres de l'équipe GIIM, qu'ils soient encore ici ou non, pour m'avoir permis de m'épanouir au quotidien dans mon travail. En particulier, Cécile Militon et Anne Moné qui m'ont fait faire mes premiers pas au laboratoire et m'ont toujours témoigné une grande amitié, à Brigitte Chebance pour sa joie de vivre, sa grande disponibilité et son accueil, à Olivier Gonçalves pour son humour décapant et sa simplicité, à Isabelle Pinto pour sa sympathie et ses délicieux cafés matinaux qui m'ont permis de démarrer agréablement la plupart de mes journées, à Sébastien Rimour pour son aide et ses conseils, à Sébastien Terrat et Jérôme Brunellière pour les moments que nous avons partagés durant nos années de thèse communes, à Ourdia Bouzid, Emilie Dumas et Sophie Comtet qui ont su me supporter dans leur bureau entre plusieurs mois et plusieurs années. Je n'oublie pas non plus Jérémie Denonfoux, dit « loulou », qui a rendu ces deux dernières années de thèse tellement plus agréables, merci pour ton amitié et pour l'entraide que nous avons su instaurer entre nous. Merci également à Nicolas Parisot, pour le vent de frais que tu apportes depuis peu à l'équipe et pour tout ce que tu as fait pour moi. Enfin, je souhaite remercier Mohieddine Missaoui et Faouzi Jaziri pour l'ensemble des développements informatiques qu'ils ont pu réaliser et qui ont fortement contribué aux travaux présentés dans cette thèse.

Je profite également de ces quelques lignes pour adresser mes sincères remerciements aux nombreux stagiaires que j'ai eu le plaisir d'encadrer, Gaëtan Guillaume, Mathieu Roudel, Anne-Sophie Yvroud, Stéphane Freitas et Laura Dumas ainsi que tous les autres stagiaires passés dans l'équipe : Florent, Mélanie, Amandine, Thomas, Maxime, Sylvain, Valérie, Lauriane, Sarah, Kévin... Un grand merci également à tous les thésards et post-docs du labo (passés et présents) avec qui nous avons passé de très bon moments, en particulier Benjamin Misson (qui partage mon quotidien depuis le lycée), Mylène Hugoni (je ne dis rien à son sujet, les gens en savent déjà bien assez...), Guillaume Borrel (pour qui la délivrance aura lieu le même jour que moi), Olivier Crouzet (pour ces Grandes Chiches inoubliables), Julie

Aufauvre, Cyril Vidau, Stéphanie Palesse, Eléonore Attard, Marion Sabart, Olivier Brouard, Marlène Jobard, Séréna Rasconi et tous les autres...

Merci au laboratoire Ampère de l'Ecole Centrale de Lyon pour m'avoir accueilli durant mon stage de M1 et mes différents passages en cours de thèse, notamment à Sandrine Demanèche, Maude David, Delina Lyon et Pascal Simonet. Merci au personnel de SITA Remédiation, en particulier à Cyrille Curvers pour sa gentillesse et sa disponibilité. Merci à Jonathan Colombet pour m'avoir accompagné et suivi lors de mes quelques sorties terrain sur le lac Pavin. Merci à l'INRA de Crouël pour m'avoir permis de faire certaines hybridations de biopuces ADN. Merci à tous les membres de l'équipe d'accueil CIDAM pour m'avoir accueilli lors de vos réunions ces six derniers mois.

Ensuite, mes remerciements vont à ma famille, mes parents et mon frère en tout premier lieu pour leur soutien et leur compréhension, mes deux grand-mères, mon Papy parti l'an dernier, ma tante Léa qui suit avec une grande attention l'avancée de mes travaux, Christian et Michelle qui m'accueillent régulièrement au sein de leur foyer et Gregory et Emilie qui ont mis au monde mon premier neveu, quel bonheur de vous compter dans mon entourage. Merci également à tous mes amis, Ben et Emilie, Bruno et Priscilla, Mathias, Amélie, Gaëtan, Aurélie, Alex, Baptiste, Mailys, Micka, Zab et Mathieu, Marie, Mary, Nath et tous ceux que j'oublie...

Pour finir, je voudrais te dédier cette thèse, Élo, pour tous les sacrifices que tu as fait pour moi, pour l'amour que tu me témoignes au quotidien mais surtout pour l'avenir que je souhaite profondément construire avec toi.

A nous...

SOMMAIRE

INTRODUCTION GENERALE	1 -
PARTIE I : SYNTHESE BIBLIOGRAPHIQUE.....	5 -
1. Les composés chlorés	6 -
1.1. Diversité et origine des molécules organohalogénées.....	6 -
1.2. Les chloroéthènes et leurs utilisations.....	8 -
1.2.1. Les chloroéthènes	8 -
1.2.2. Synthèse des chloroéthènes	8 -
1.2.3. Les productions industrielles.....	9 -
1.3. Propriétés des chloroéthènes et devenir dans l'environnement.....	11 -
1.3.1. Propriétés.....	11 -
1.3.2. Devenir des solvants chlorés	11 -
1.4. Risques sanitaires	13 -
2. Biodégradation des chloroéthènes	14 -
2.1. Le co-métabolisme aérobie	14 -
2.2. L'assimilation aérobie	16 -
2.3. Déchloration oxydative anaérobie	17 -
2.4. La déchloration réductrice anaérobie	18 -
2.4.1. Déchloration réductrice métabolique : la déhalorespiration.....	19 -
a) Eléments intervenant dans la déhalorespiration	19 -
b) Diversité des microorganismes déhalorespirants et spécificité de leurs enzymes	20 -
c) Conditions physico-chimiques et biologiques propices à la déhalorespiration	23 -
2.4.2. Déchloration réductrice co-métabolique	24 -
3. Les Dehalococcoides	26 -
3.1. Caractéristiques des souches affiliées au genre <i>Dehalococcoides</i>	26 -
3.1.1. Caractéristiques morphologiques et physiologiques	27 -
3.1.2. Caractéristiques génomiques.....	27 -
3.1.3. Transferts horizontaux de gènes chez <i>Dehalococcoides</i>	29 -
3.2. Les autres groupes déhalorespirants affiliés aux <i>Chloroflexi</i>	30 -
3.2.1. Le genre <i>Dehalogenimonas</i>	30 -
3.2.2. Le groupe des <i>Dehalococcoides</i> -like (DLG).....	30 -
4. Stratégies de réhabilitation des aquifères contaminées par les solvants chlorés ..	32 -
4.1. Approches physiques.....	33 -
4.2. Approches chimiques	35 -
4.3. Approches biologiques	36 -
4.3.1. La phytoremédiation	37 -
4.3.2. La bioremédiation microbienne	39 -
a) L'atténuation naturelle	40 -
b) La biostimulation	41 -
c) La bioaugmentation	43 -
4.4. Futures directions	44 -
5. Les apports des outils « haut débit » pour le développement de stratégies de bioremédiation	46 -
5.1. Révolution des approches moléculaires et émergence des outils haut débit.....	48 -
5.1.1. Les méthodes basées sur l'amplification PCR	48 -
5.1.2. Les approches haut débit dites de méta « omiques »	50 -
5.2. Les biopuces ADN	53 -
5.2.1. Les biopuces phylogénétiques (POA)	54 -
5.2.2. Les biopuces fonctionnelles (FGA).....	56 -
Conclusion Générale	59 -

PARTIE II : DETECTION DE SEQUENCES INCONNUES A L'AIDE DES BIOPUCES ADN : STRATEGIES DE DESIGN DE SONDAS EXPLORATOIRES.- 61 -

1. Contexte.....	62 -
2. Objectifs	62 -
ARTICLE 1: Detecting unknown sequences with DNA microarrays: explorative probe design strategies	64 -
3. Discussion	89 -

PARTIE III : DEVELOPPEMENT D'UN LOGICIEL DE SELECTION DE SONDAS OLIGONUCLEOTIDIQUES POUR BIOPUCES ADN FONCTIONNELLES.....- 91 -

1. Contexte.....	92 -
2. Objectifs	93 -
3. Principaux résultats	93 -
ARTICLE 2: HiSpOD: probe design for functional DNA microarrays	96 -
4. Discussion	118 -

PARTIE IV : ETUDE *IN SITU* DES POPULATIONS DEHALORESPIRANTES IMPLIQUEES DANS LA DEGRADATION DU TCE SUR DES SITES EN COURS DE REHABILITATION PAR DES APPROCHES DE BIOSTIMULATION.....- 120 -

1. Contexte.....	121 -
2. Objectifs	122 -
3. Principaux résultats	123 -
ARTICLE 3: Complete TCE degradation to ethene mediated by a complex dehalorespiring community during an <i>in situ</i> biostimulation process	125 -
4. Discussion	148 -

CONCLUSION ET PERSPECTIVES 151 -

1. Conclusion	152 -
2. Perspectives	155 -
2.1. Amélioration des outils de sélection de sondes	155 -
2.2. Généralisation de l'utilisation de nos logiciels	157 -
2.3. Etude des processus de déhalorespiration dans les milieux naturels : identification de nouvelles souches et de nouvelles capacités métaboliques	158 -
2.3.1. Les milieux d'eau douce, sources de nouvelles capacités métaboliques	159 -
2.3.2. Identification de nouvelles potentialités métaboliques : intérêts biotechnologiques.....	160 -

REFERENCES 162 -

ANNEXES

LISTE DES FIGURES

Figure 1 : Formules chimiques des composés de la famille des chloroéthènes.

Figure 2 : Réactions chimiques utilisées pour synthétiser les chloroéthènes.

Figure 3 : Migration des chloroéthènes dans le milieu souterrain suite à une contamination d'origine industrielle.

Figure 4 : Voie de biodégradation du trichloroéthylène (TCE) par co-métabolisme aérobie chez les bactéries méthanotrophes.

Figure 5 : Mécanisme d'assimilation aérobie du chlorure de vinyle (VC) empruntant la voie classique de dégradation de l'éthylène.

Figure 6 : Mécanisme supposé d'assimilation aérobie du dichloroéthylène (DCE) par la souche *Polaromonas chloroethenica* JS666.

Figure 7 : Déchloration réductrice anaérobie des chloroéthènes par différentes souches procaryotiques.

Figure 8 : Organisation des opérons contenant les gènes codant pour les déhalogénases réductrices chez différentes souches déhalorespirantes.

Figure 9 : Mécanisme réactionnel de la déhalorespiration.

Figure 10 : Potentiels redox caractérisant les principaux processus respiratoires identifiés chez les microorganismes.

Figure 11 : Succession des différents processus respiratoires le long du panache dans un aquifère pollué par des chloroéthènes

Figure 12 : Arbre phylogénétique représentant les huit groupes du phylum des *Chloroflexi*, soit les sept sous-phyla déjà décrits et le groupe supplémentaire des *Dehalococcoides*-like (DLG).

Figure 13 : Clichés de microscopie électronique de plusieurs souches de *Dehalococcoides*.

Figure 14 : Comparaison des génomes de cinq souches de *Dehalococcoides* (souches VS, 195, BAV1, CBDB1 et GT) réalisée à l'aide du logiciel GView en utilisant comme référence le génome de la souche VS.

Figure 15 : Premières pages d'articles parus dans les années 1990 dans la revue Environmental Science and Technology.

Figure 16 : Représentation schématique des principales contraintes rencontrées pour la mise en place de stratégies de remédiation efficaces après pollution d'une nappe phréatique par les chloroéthènes.

Figure 17 : Principe de la technique d'air sparging.

Figure 18 : Utilisation de surfactants ou d'alcools pour traiter les aquifères contaminés par des chloroéthènes.

Figure 19 : Réactions chimiques permettant l'oxydation du TCE.

Figure 20 : Technique de remédiation chimique par injection directe d'une solution oxydante dans la nappe phréatique.

Figure 21 : Principe des barrières réactives passives placées en aval de la zone source de contamination.

Figure 22 : Représentation schématique des différentes méthodes de phytoremédiation.

Figure 23 : Principe de l'Enhanced Reductive Dechlorination (ERD) par injection directe de substrat dans l'aquifère (A) ou par la technique de re-circulation (B).

Figure 24 : Dispositif d'électro-stimulation séquentielle anaérobie/aérobie mis en place sur un site contaminé par du PCE.

Figure 25 : Caractérisation du microbiote oral humain par la technique FISH grâce à l'utilisation d'un mélange de sondes ciblant 15 *taxa* (genres ou familles) classiquement retrouvés dans la cavité buccale.

Figure 26 : Analyse FISH-nanoSIMS de l'incorporation d'azote par le *consortia* ANME-2/DSS incubé avec du $^{15}\text{N}_2$.

Figure 27 : Principe des technologies de séquençage de troisième génération.

Figure 28 : Principe des biopuces ADN.

Figure 29 : Représentation graphique de l'évolution du nombre de nucléotides présents dans la base de données EMBL pour la période comprise entre 1982 et 2010

Figure 30 : Photographie aérienne du lac Pavin (Puy de Dôme, France).

Figure 31 : Arbre phylogénétique représentant les huit groupes du phylum des *Chloroflexi* et intégrant les séquences isolées de la zone anoxique du lac Pavin.

LISTE DES TABLEAUX

Tableau 1 : Familles de composés organochlorés les plus utilisés ou rejetés dans l'environnement.

Tableau 2 : Propriétés physico-chimiques des chloroéthènes.

Tableau 3 : Temps de demi-vie des différents chloroéthènes observés en fonction des processus de dégradation et concentrations seuils fixées par les autorités sanitaires.

Tableau 4 : Processus microbiens permettant la biodégradation des chloroéthènes.

Tableau 5 : Donneurs et accepteurs d'électrons utilisés par les principaux microorganismes déhalorespirants.

Tableau 6 : Activités des différentes déhalogénases réductrices identifiées chez des bactéries déhalorespirantes.

Tableau 7 : Concentrations seuils en H_2 nécessaires aux groupes microbiens hydrogénotrophes en fonction de leur métabolisme.

Tableau 8 : Liste des molécules organohalogénées dégradées par les différentes souches de *Dehalococcoides* et des métabolites qu'elles produisent au cours du processus de déhalorespiration.

Tableau 9 : Taille et composition des génomes des souches déhalorespirantes obligatoires appartenant aux genres *Dehalococcoides* et *Dehalogenimonas*.

Tableau 10 : Répartition approximative des volumes d'eau douce dans les différents compartiments terrestres.

Tableau 11 : Réactions de fermentation de molécules carbonées simples conduisant à la production d' H_2 .

Tableau 12 : Comparaison des points forts et des principales limites des différentes plateformes de séquençage.

Tableau 13: Comparaison des approches de méta«omiques» et de biopuces ADN pour des applications en écologie microbienne.

LISTE DES ABREVIATIONS

%GC	Pourcentage en Guanine et Cytosine	IUPAC	International Union of Pure and Applied Chemistry
16S	16 Svedberg	Koc	Coefficient d'adsorption
ADEME	Agence De l'Environnement et de la Maîtrise de l'Energie	Kpb	Kilo paire de bases
ADN	Acide DésoxyriboNucléique	Kb	Kilo bases
ADNc	ADN complémentaire	MAR-FISH	MicroAutoRadiography-FISH
ADNg	ADN génomique	MMO	Méthane MonoOxygénase
ADNr	ADN ribosomique	Mpb	Méga paire de bases
AkMO	Alcène monooxygénase	NanoSIMS	Spectromètre de masse à ionisation secondaire à l'échelle nanométrique
ANR	Agence Nationale de la Recherche	NGS	Next Generation Sequencing
ARDRA	Amplified Ribosomal DNA Restriction Analysis	OGM	Organisme Génétiquement Modifié
ARN	Acide RiboNucléique	OMS	Organisation Mondiale de la Santé
ARNa	ARN antisens	ORF	Open Reading Frame (cadre de lecture ouvert)
ARNm	ARN messenger	ori	origine de réplication
ARNr	ARN ribosomique	OTU	Operational Taxonomic Unit
ATCC	American Type Culture Collection	Pa	Pression de vapeur
ATP	Adénosine TriPhosphate	PBDE	Ethers diphényles polybromés
BLAST	Basic Local Alignment Search Tool	PCA	Tétrachloroéthane
CAD	ChloroAcétaldéhyde Déshydrogénase	PCB	PolyChloroBiphényle
CARD-FISH	CAtalysed Reporter Deposition-FISH	PCDD/F	Dibenzo-p-dioxine et furane polychloré
CDS	Coding DNA Sequence	PCE	Tétrachloroéthylène
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act	PCN	Naphtalène polychloré
CFC	ChloroFluoroCarbure	PCR	Polymerase Chain Reaction
CIRC	Centre International de Recherche sur le Cancer	PeCA	PentaChloroéthAne
CNRS	Centre National de la Recherche Scientifique	POA	Plylogenetic Oligonucleotide Array
COV	Composé Organique Volatil	ppm	partie par million
COHV	Composé OrganoHalogéné Volatil	PVC	Polychlorure de vinyle
CPU	Central Processing Unit	qPCR	PCR quantitative
DCA	DiChloroéthAne	RACE-PCR	Rapid Amplification of cDNA-ends by PCR
DCE	DiChloroéthylEne	Rdh	Déhalogénase réductrice
DDT	DichloroDiphénylTrichloroéthane	RDP	Ribosomal Database Project
DGGE	Denaturing Gradient Gel Electrophoresis	RISA	Ribosomal Intergenic Spacer Analysis
DLG	Dehalococcoides-Like Group	RT-PCR	Reverse Transcription-PCR
DNAPL	Dense Non Aqueous Phase Liquid	SDS	Sodium DodecylSulfate
dNTP	désoxyNucléotide Tri-Phosphate	SIMSISH	Secondary Ion Mass Spectrometry <i>In Situ</i> Hybridization
EaCoMT	Epoxyalcane Coenzyme M Transférase	SIP	Stable Isotope Probing
EBI	European Bioinformatics Institute	SMRT	Single Molecule Real Time Technology
EDTA	Ethylène Diamine Tétra Acétique	SNR	Signal to Noise Ratio
EGEE	Enabling Grid for E-sciences in Europe	SSCP	Single-Strand Conformation Polymorphism
EMBL	European Molecular Biology Laboratory	SVE	Système d'extraction de vapeurs
EPA	Environmental Protection Agency	TCA	TriChloroéthAne
ERC	European Research Council	TCE	TriChloroéthylEne
ERD	Enhanced Reductive Dechlorination	TGGE	Temperature Gradient Gel Electrophoresis
FGA	Functional Gene Array	T_m	Melting Temperature (ou température de fusion)
FISH	Fluorescent In Situ Hybridization	TDO	Toluène DiOxygénase
FRET	Fluorescence Resonance Energy Transfer	TMO	Toluène MonoOxygénase
GOLD	Genome OnLine Database	T-RFLP	Terminal Restriction Fragment Length Polymorphism
GPU	Graphics Processing Unit	UTR	Région non Transcrite
GSH	Glutathion	UV	Ultra Violet
GST	Glutathion-S-Transférase	VC	Chlorure de vinyle
HAD	HaloAcide Déhalogénase	WGA	Whole Genome Array
HAP	Hydrocarbure Aromatique Polycyclique		
HPR	High Plasticity Regions		
ITRC	Interstate Technology and Regulatory Council		

INTRODUCTION GENERALE

L'industrialisation de notre société, qui s'est accélérée au cours du XX^{ème} siècle, a conduit à la production d'une grande diversité de molécules chimiques qui, après leur rejet dans l'environnement, ont engendré la pollution de nombreux écosystèmes. Pour la France, la base de données BASOL (<http://basol.ecologie.gouv.fr/>) recense actuellement 4455 sites pollués nécessitant une action des pouvoirs publics à titre préventif ou curatif. Près de 17% de ces sites sont contaminés par des solvants halogénés, parmi lesquels figurent les molécules de la famille des chloroéthènes, polluants majeurs des eaux souterraines et des nappes phréatiques. Or, ces compartiments, considérés comme extrêmement fragiles, représentent la principale source d'eau douce, élément indispensable à la vie terrestre. Leur protection est donc devenue un objectif majeur. Cette prise de conscience s'est traduite, à partir des années 1980, par l'adoption de textes réglementaires dont l'objectif était de prévenir les pollutions industrielles.

Dans le même temps, une politique de lutte, de prévention et de surveillance des substances polluantes a été mise en place dans le but de limiter, voire de supprimer, leur impact sur les eaux souterraines. Alliant les connaissances acquises dans les domaines de la chimie, de la physique et de la biologie, des équipes pluridisciplinaires ont développé une large gamme de procédés adaptés à la restauration (également appelée remédiation) des nappes phréatiques contaminées par les chloroéthènes (Henry *et al.*, 2002). Néanmoins, depuis quelques années, un intérêt plus particulier est porté aux approches biologiques car elles sont peu invasives, faciles à mettre en œuvre, peu coûteuses et très bien acceptées par l'opinion publique. Ces approches, dites de « bioremédiation », exploitent les capacités naturelles de certains organismes, comme les végétaux (phytoremédiation) ou les microorganismes (bioremédiation microbienne), à pouvoir dégrader les substances polluantes (Perelo, 2010). Pour les chloroéthènes, l'essentiel des approches de bioremédiation se base sur les capacités métaboliques portées par les microorganismes, la phytoremédiation restant relativement marginale. Pour certains sites pollués, la seule présence de certaines souches adaptées est suffisante pour dégrader les polluants présents sans intervention humaine. C'est ce que l'on appelle l'atténuation naturelle.

La biodégradation des chloroéthènes peut avoir lieu aussi bien en présence qu'en absence d'oxygène par l'intermédiaire de plusieurs voies métaboliques (Mattes *et al.*, 2010; Pant and Pant, 2010). La voie la plus efficace est la déchloration réductrice qui se déroule en anaérobiose et qui nécessite l'intervention de souches affiliées au genre *Dehalococcoides* pour la réalisation des dernières étapes de la voie métabolique conduisant à l'éthylène, molécule non toxique. Sur la base de ces connaissances, différentes stratégies de bioremédiation visant

à accroître l'efficacité du processus d'atténuation naturelle peuvent être envisagées : la biostimulation grâce à l'ajout de substrats et la bioaugmentation grâce à l'apport de microorganismes épurateurs directement sur le site contaminé (Tyagi *et al.*, 2011).

Une bonne connaissance de la microflore indigène présente *in situ* ainsi que de ses capacités métaboliques s'avère aujourd'hui indispensable pour la mise en œuvre de stratégies de bioremédiation efficaces. Or, l'acquisition de ces données représente encore un challenge tant les communautés microbiennes présentes dans les écosystèmes environnementaux sont denses et diversifiées (Torsvik *et al.*, 1990; Whitman *et al.*, 1998). Cependant, grâce à l'essor des approches moléculaires et l'avènement d'outils qualifiés de « haut débit », notre vision du monde microbien a complètement changé. Il est désormais possible de caractériser la majorité des populations microbiennes à partir d'un nombre quasi-illimité de biomarqueurs (phylogénétiques et fonctionnels). Les biopuces ADN fonctionnelles sont parmi les outils les plus pertinents pour évaluer la diversité fonctionnelle des écosystèmes perturbés du fait de leur simplicité d'utilisation, de la rapidité d'acquisition de données pour un grand nombre d'échantillons et de leur facilité d'interprétation. Les industriels spécialisés dans la dépollution se tournent d'ailleurs de plus en plus vers l'utilisation de cette technologie, jusqu'à présent principalement exploitée par le domaine médical, pour le développement d'outils diagnostics performants. Cependant, les stratégies de détermination de sondes adaptées à la problématique environnementale sont différentes et nécessitent le développement de nouveaux logiciels.

Afin de bien cerner le contexte de ce travail, ses enjeux et ses principaux résultats, ce mémoire, rédigé sous la forme d'une thèse sur articles, s'articule autour de quatre parties :

- **La première partie (Partie I)** est consacrée à la synthèse bibliographique et a pour objectif de replacer le contexte scientifique de la thèse. Elle se découpe en cinq chapitres, chacun centré sur un des principaux domaines de recherche en relation avec les articles présentés dans les parties suivantes. Le premier chapitre est consacré à la description des composés chlorés, et plus particulièrement des chloroéthènes, qui sont à la base de ce travail. Les chapitres 2 et 3 sont dédiés à la description des voies métaboliques empruntées par les microorganismes pour la biodégradation de ces molécules et d'un groupe bactérien particulièrement bien adapté, les *Dehalococcoides*. En effet, il s'agit du seul groupe capable de dégrader l'ensemble des chloroéthènes. Dans le quatrième chapitre, sont décrites toutes les techniques de remédiation utilisées pour la décontamination d'aquifères pollués par les chloroéthènes, ce qui permet de souligner les avantages des approches de bioremédiation. Pour conclure cette partie, le dernier chapitre fait un état des connaissances sur les outils

moléculaires qualifiés de « haut débit » développés au cours des dernières années, en insistant plus particulièrement sur la technologie biopuce ADN pour mettre en évidence son intérêt pour aider à la mise en place de stratégies de bioremédiation efficaces.

- **Dans la partie II**, toutes les stratégies de sélection de sondes pour biopuces ADN développées ces dernières années sont synthétisées sous la forme d'une revue scientifique. Ce travail est volontairement séparé de la partie bibliographique afin de pouvoir présenter plus en détails ces approches et notamment, les stratégies innovantes de détermination de sondes exploratoires élaborées au sein de notre équipe de recherche (Militon *et al.*, 2007 ; Terrat *et al.*, 2010 : ARTICLE ANNEXE 1 ; Dugat-Bony *et al.*, 2011 : ARTICLE 2 PARTIE III). Ces sondes permettent d'accéder aux séquences de gènes connues mais aussi à celles encore inconnues, critère crucial pour une étude approfondie des populations microbiennes environnementales.

- **Les parties III et IV** présentent, sous la forme de deux articles, les principaux résultats issus directement du travail de thèse. Dans la partie III, le logiciel de design de sondes pour biopuces ADN fonctionnelles mis au point, ainsi que sa validation, sont présentés. Cet outil, nommé HiSpOD, intègre des concepts innovants permettant la sélection de sondes exploratoires, très sensibles et très spécifiques, notamment grâce à la construction et à l'incorporation d'une base de données représentative de l'ensemble des gènes potentiellement retrouvés dans l'environnement. La validation expérimentale des sondes sélectionnées a été réalisée avec la construction et l'hybridation d'une biopuce ADN fonctionnelle dédiée aux voies de biodégradation des chloroéthènes, appelée DechloArray. Dans la partie IV, l'utilisation de cette biopuce est étendue à la caractérisation d'un plus grand nombre d'échantillons environnementaux prélevés directement sur des sites contaminés. Ce travail, réalisé dans le cadre d'un programme ANR PRECODO (Evasol), a confirmé la pertinence de notre outil DechloArray et son intérêt pour les industriels pour le diagnostic des capacités de biodégradation des microflore présentes sur les sites pollués et l'aide à la décision dans la mise en place des traitements de bioremédiation.

Les différents points abordés au cours de ce mémoire sont enfin synthétisés dans une conclusion générale. Les principales perspectives de recherche sont également indiquées afin d'aller plus loin dans la compréhension des mécanismes impliqués dans les processus de dégradation de molécules polluantes.

PARTIE I : SYNTHESE BIBLIOGRAPHIQUE

1. Les composés chlorés

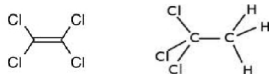
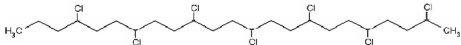
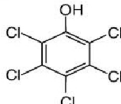
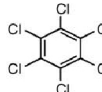
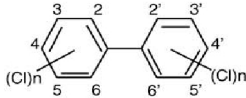
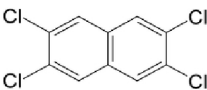
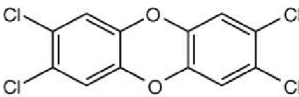
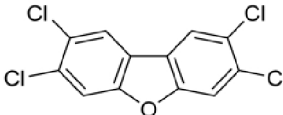
Le chlore (Cl) est l'un des éléments chimiques les plus abondants sur notre planète (Oberg, 2002). Avec le fluor (F), le brome (Br) et l'iode (I), ils forment le groupe des halogènes. Ces éléments sont présents en grande quantité dans les océans et la croûte terrestre sous forme de sels et entrent également dans la composition de nombreuses molécules organiques alors appelées organohalogènes. Ces molécules, apparentées aux hydrocarbures, ont longtemps été considérées comme étant principalement d'origine anthropique. En effet, au cours des dernières décennies, l'Homme a synthétisé de nombreuses molécules organohalogénées. Ensuite, leurs utilisations et leurs rejets, le plus souvent accidentels, ont eu des conséquences extrêmement néfastes pour l'environnement. Cependant, certains travaux menés dans le domaine de la chimie laissent penser que les processus naturels pourraient également produire non seulement de grandes quantités (le plus souvent lors des réactions épisodiques comme les feux de forêt ou les éruptions volcaniques) mais aussi une large diversité de molécules organohalogénées (Gribble, 1994).

1.1. Diversité et origine des molécules organohalogénées

Près de 4000 molécules naturelles halogénées ont récemment été recensées et depuis leur nombre ne cesse de croître à un rythme d'environ 100 à 200 nouvelles molécules par an (Gribble, 2004). Leurs structures sont très diverses avec des molécules de type aliphatiques, alicycliques et aromatiques (mono- ou poly-), celles-ci pouvant être substituées par un ou plusieurs groupements halogènes. Ces molécules peuvent être issues de processus très différents :

- **Des processus biotiques :** la quasi totalité des organismes vivants uni- et pluricellulaires synthétise des composés organohalogénés nécessaires généralement à leurs mécanismes de défenses ou à leurs développements cellulaires. La plus grande diversité de ces molécules est générée par les organismes marins (Gribble, 2003). Cependant de très nombreux organismes terrestres tels que les plantes supérieures, les champignons, les lichens, les insectes et les animaux supérieurs sont également producteurs de telles molécules. A titre d'exemple, il a été montré que les algues retrouvées dans les environnements marins sont à l'origine de l'émission de quantités importantes d'organohalogènes volatils tels que les chloroéthènes (Abrahamsson *et al.*, 1995), le chlorométhane, le bromoforme et le bromoéthane (Laturnus, 2001), qui sont le plus souvent utilisés dans leurs mécanismes de

Tableau 1 : Familles de composés organochlorés les plus utilisés ou rejetés dans l’environnement.

Famille	Formule type	Utilisations principales ou sources de rejet
Solvants chlorés (exemples: PCE et 1,1,1-TCA)		Solvants, dégraissage des métaux, lavage à sec
Chlorofluorocarbures (CFCs)	$F_x - C_n - Cl_y$	Fluides frigorigènes, aérosols
Paraffines chlorées		Huiles de coupe, retardateurs de flamme, additifs
Chlorophénols (exemple: pentachlorophénol)		Biocides (fongicides, bactéricides), intermédiaires de synthèse
Chlorobenzènes (exemple: hexachlorobenzène)		Intermédiaires de synthèse
Biphényles polychlorés (PCBs)		Fluides de transformateurs électriques, fluides hydrauliques, additifs
Naphtalènes polychlorés (PCN)		Fluides de condensateurs, isolants électriques, lubrifiants (production arrêtée depuis les années 1980s)
Dioxines chlorées (PCDDs)		Combustion, incinération des déchets
Dibenzofuranes chlorés (PCDFs)		Combustion, incinération des déchets

défense ou en réponse à différents stress environnementaux (Bravo-Linares *et al.*, 2010). Parmi l'ensemble des molécules produites, certaines présentent des intérêts considérables notamment pour des applications dans le domaine de la médecine comme par exemple les antibiotiques et plus particulièrement la vancomycine. Cet antibiotique, de la famille des glycopeptides, a été isolé à partir de la souche bactérienne *Amycolatopsis orientalis* et représente l'un des derniers remparts contre les maladies nosocomiales (Nicolaou *et al.*, 1999; Gao, 2002). Nous pouvons également citer comme exemple l'épibatidine, composé initialement isolé de la peau d'une grenouille équatoriale et qui possède une activité analgésique plus puissante que la morphine (Spande *et al.*, 1992). Même si cette molécule n'est pas directement appliquée en médecine du fait de sa forte toxicité, l'utilisation d'analogues pourrait être envisagée.

- **Des processus abiotiques** : les molécules organohalogénées sont produites au cours de réactions géothermiques telles que les éruptions volcaniques, les processus de combustion de biomasse comme les incendies de forêts, ou encore les processus de formation des sols et des sédiments du fait notamment de la décomposition de la matière organique. Pour exemple, parmi les 134 molécules halogénées différentes détectées dans les fumeroles et les gaz de laves de volcans en activité, sont retrouvées des haloalcanes (chloro/bromo/fluoro/iodo-méthanes, chloro/bromo/fluoro/iodo-éthanes et chloropropane), des haloalcènes (chloro/bromo/fluoro/iodo-éthènes et chloro/bromo/fluoro-propènes), des halobenzènes et des halofuranes (Jordan *et al.*, 2000; Jordan, 2003).

Parallèlement à l'émission naturelle, l'industrie chimique est également responsable de rejets importants dans l'environnement d'une grande diversité de composés organohalogénés synthétiques. En effet, ils sont utilisés pour l'élaboration de très nombreux composés d'usage courant. Le chlore est le principal halogène utilisé avec plus de 60% de l'industrie chimique européenne portant sur la production de molécules chlorées ou employant cet halogène pour l'élaboration de molécules d'intérêt (Contreras Lopez, 2003). Les composés organochlorés les plus utilisés ou rejetés dans l'environnement sont les solvants chlorés, les chlorofluorocarbures (CFCs), les paraffines chlorées, les biphényles polychlorés (PCBs), les chlorophénols, les dioxines et les dibenzofuranes chlorés, les naphthalènes polychlorés (PCN) ou encore les chlorobenzènes (Tableau 1) (Vandecasteele, 2005).

Dans le cadre de ce travail, nous nous sommes intéressés à la famille des solvants chlorés et plus particulièrement aux chloroéthènes qui sont considérés comme les polluants

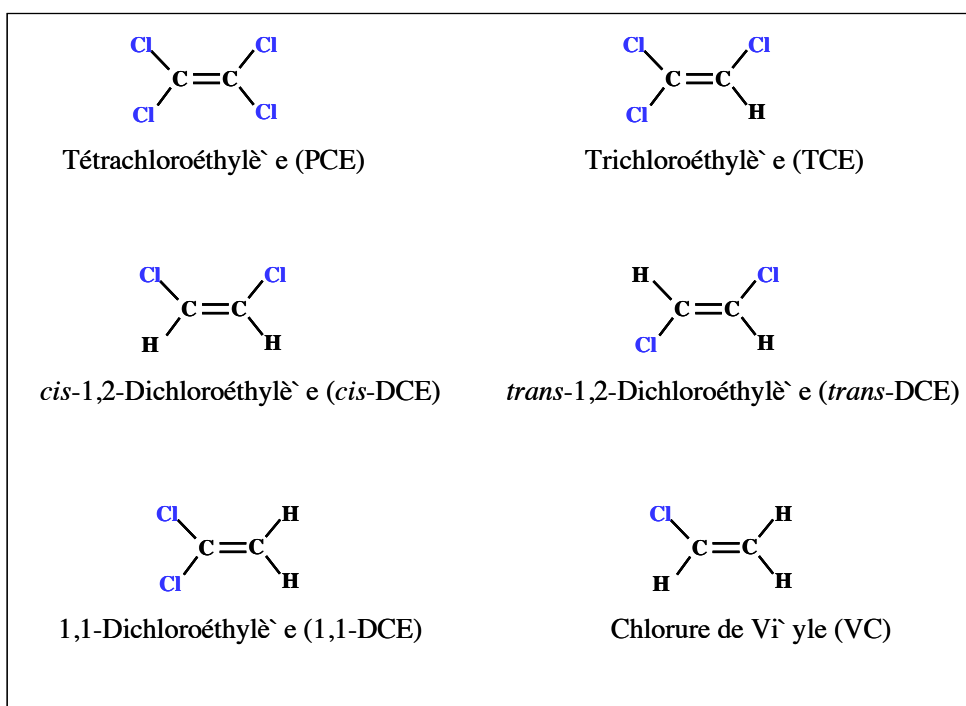


Figure 1 : Formules chimiques des composés de la famille des chloroéthènes.

majeurs des eaux souterraines et des nappes phréatiques (McCarty, 1997). Contrairement à d'autres composés chlorés, ils ne peuvent pas être assimilés à des xénobiotiques au sens strict du terme car ils sont également produits naturellement lors de certaines réactions biochimiques ou géochimiques. En effet, ils ont été détectés dans l'eau de mer, dans des zones dépourvues d'activité humaine mais aussi dans des gaz issus d'éruptions volcaniques (Abrahamsson *et al.*, 1995; Jordan *et al.*, 2000).

1.2. Les chloroéthènes et leurs utilisations

1.2.1. Les chloroéthènes

Les chloroéthènes sont des hydrocarbures aliphatiques. Avec les chlorométhane (chlorure de méthyle, chlorure de méthylène, chloroforme et tétrachlorure de carbone) et le 1,1,1-trichloroéthane (1,1,1,-TCA), ils forment le groupe des solvants chlorés. Leur squelette carboné se compose de deux atomes de carbone reliés par une double liaison covalente (éthylène de formule brute C_2H_4). Chacun présente une ou plusieurs substitution(s) par des atomes de chlore. Cette famille contient, dans l'ordre décroissant du nombre de substitutions, le tétrachloroéthylène (perchloroéthylène, PCE), le trichloroéthylène (TCE), le dichloroéthylène (DCE) et le chlorure de vinyle (VC) (Figure 1). Parmi ces molécules, seul le DCE a plusieurs formes qui se distinguent par la position des deux atomes de chlore sur le squelette carboné. Ainsi, pour le 1,1-DCE, ils se trouvent sur le même carbone alors que pour le 1,2-DCE, ils sont portés par deux carbones différents. Par ailleurs, il existe deux isomères différents pour le 1,2-DCE, appelés *cis* et *trans*. Pour l'isomère *cis*, les 2 atomes de chlore se trouvent du même côté de l'axe de la liaison $C=C$ alors que pour le *trans*, ils sont de part et d'autre de cet axe.

1.2.2. Synthèse des chloroéthènes

La synthèse des chloroéthènes est réalisée grâce à un ensemble de réactions chimiques. Ainsi, le PCE est obtenu par chloration ou oxychloration du 1,2-dichloroéthane (1,2-DCA) (Figure 2A), par craquage d'un mélange de tétrachloroéthane (PCA) et de pentachloroéthane (PeCA), ou encore par chloration de résidus chlorés tels que le VC (<http://www.societechimiquedefrance.fr>). Les deux premières réactions aboutissent également à la production de TCE. En plus des réactions citées précédemment, le TCE est également

produit à partir de l'acétylène, par chloration et formation de PCA symétrique ($\text{CHCl}_2\text{-CHCl}_2$) qui est ensuite craqué sur charbon actif (Figure 2B). Sa production peut également se faire par chloration successive de l'éthylène avec formation de 1,2-DCA, puis de 1,1,2-trichloroéthane (1,1,2-TCA) et de PCA qui sont ensuite craqués pour donner du TCE. Enfin, il peut être formé par hydrogénation du PCE (<http://www.societechimiquedefrance.fr>).

Le 1,2-DCE commercial est un mélange des deux isomères *cis* et *trans*, avec prédominance de l'isomère *cis*. Sa synthèse se fait par chloration directe de l'acétylène en présence d'un catalyseur. Cependant, il peut également être recueilli comme sous-produit lors de la synthèse de composés comme le VC et le TCE. Le craquage thermique du 1,1,2-TCA donne, quant à lui, naissance à la fois à du 1,1 et du 1,2-DCE dont les proportions peuvent être ajustées grâce à la température ou sous l'effet de catalyseurs (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – 1,2-Dichloroéthylène, 2005). Le 1,1-DCE résulte principalement de la réaction du 1,1,2-TCA avec de l'hydroxyde de sodium ou de calcium. Cependant, la décomposition thermique de méthyl chloroforme (1,1,1-TCA) à plus de 350°C, ou à des températures inférieures à 180°C en présence de cuivre, génère aussi du 1,1-DCE (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – 1,1-Dichloroéthylène, 2005).

La synthèse du VC se fait, quant à elle, en deux étapes successives à partir d'éthylène (Figure 2C). Au cours de la première étape, du 1,2-DCA est synthétisé par chloration directe de l'éthylène avec du chlore ou par oxychloration avec de l'HCl. La seconde étape aboutit au VC par une réaction de pyrolyse du 1,2-DCA en phase gazeuse (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – Chlorure de vinyle, 2010).

1.2.3. Les productions industrielles

Les chloroéthènes sont des produits ininflammables mais aussi apolaires ce qui en fait de bons solvants des graisses. Grâce à ces propriétés, ils ont été très largement utilisés dès le début du XX^{ème} siècle comme produits de dégraissage des métaux et des textiles pour de nombreuses activités industrielles.

Le PCE, qui est l'alcène chloré le plus substitué, trouve plusieurs applications dans divers secteurs de la chimie. Ainsi, bien que principalement utilisé pour le nettoyage à sec des textiles, il est également employé pour le dégraissage des pièces métalliques dans les industries automobiles et métallurgiques, comme intermédiaire de synthèse dans la fabrication d'hydrocarbures fluorés, comme décapant pour la peinture et l'encre d'imprimerie, ou encore

comme diélectrique ou isolant dans les appareils électriques tels que les transformateurs et les condensateurs (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – Tétrachloroéthylène, 2008). En 2004, ces différentes activités ont nécessité l'usage de 54 000 tonnes de PCE en Europe de l'Ouest et de 191 000 tonnes aux Etats-Unis (<http://www.societechimiquedefrance.fr>).

Le TCE est également un très bon solvant. Comme le PCE, il est surtout employé dans l'industrie automobile où il sert au dégraissage des pièces métalliques. Les autres applications concernent l'industrie textile pour le nettoyage à sec du coton et de la laine mais aussi la fabrication de divers produits comme les adhésifs, les lubrifiants, les peintures, les vernis et les pesticides. Il entre également dans les procédés de fabrication de produits pharmaceutiques, de retardateurs chimiques d'inflammation et d'insecticides (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – Trichloroéthylène, 2005). En 1998, la consommation de TCE par les Etats-Unis a été évaluée à 78 000 tonnes et elle était de 33 000 tonnes pour l'Europe de l'Ouest en 2004 (<http://www.societechimiquedefrance.fr>).

Dans le secteur de l'industrie, le 1,1-DCE sert à produire du 1,1,1-TCA. Il peut aussi facilement co-polymériser avec le VC et d'autres monomères pour former des fibres synthétiques. Les polymères ainsi formés sont utilisés comme revêtements des films d'hydrate de cellulose destinés à l'emballage des produits alimentaires (exemple : le Saran) ou de polypropylène afin d'en diminuer la perméabilité à l'eau et à l'air. Le 1,2-DCE sert, quant à lui, principalement d'intermédiaire chimique pour la synthèse de solvants chlorés comme le TCE et le PCE. Il est également utilisé comme solvant notamment pour les graisses ou les phénols, pour l'extraction à froid de substances sensibles à la température (parfum, caféine), comme réfrigérant ou encore comme agent de retardement de la fermentation. Au cours de ces réactions, l'isomère *trans* est plus largement utilisé que le *cis* ou que le mélange des deux (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – 1,2-Dichloroéthylène, 2005).

Enfin, le VC, substitué en une seule position, est largement utilisé comme monomère dans la fabrication de matières plastiques (bouteilles, revêtements de sol, PVC et co-polymères), dans les synthèses organiques mais aussi comme réfrigérant. Il trouve également de nombreuses applications dans la fabrication de produits utilisés dans le bâtiment, l'industrie automobile, l'isolation de câbles et de fils électriques, les tuyauteries, l'équipement industriel et ménager (Ineris – Fiche de données toxicologiques et environnementales des substances chimiques – Chlorure de vinyle, 2010).

Tableau 2 : Propriétés physico-chimiques des chloroéthènes.

Composé	Formule brute	Pa à 20°C	Koc (mg/g)	Solubilité dans l'eau (mg/L) à 20°C	Densité (g/cm ³) à 20°C
PCE	C ₂ Cl ₄	1900	364	150	1,63
TCE	C ₂ HCl ₃	7960	126	1100	1,46
1,1-DCE	C ₂ H ₂ Cl ₂	66500	65	2250	1,22
cis-1,2-DCE	C ₂ H ₂ Cl ₂	24000	49	800	1,28
trans-1,2-DCE	C ₂ H ₂ Cl ₂	35300	59	600	1,26
VC	C ₂ H ₃ Cl	330000	57	2670	0,91

Pa : pression de vapeur ; Koc : coefficient d'absorption.

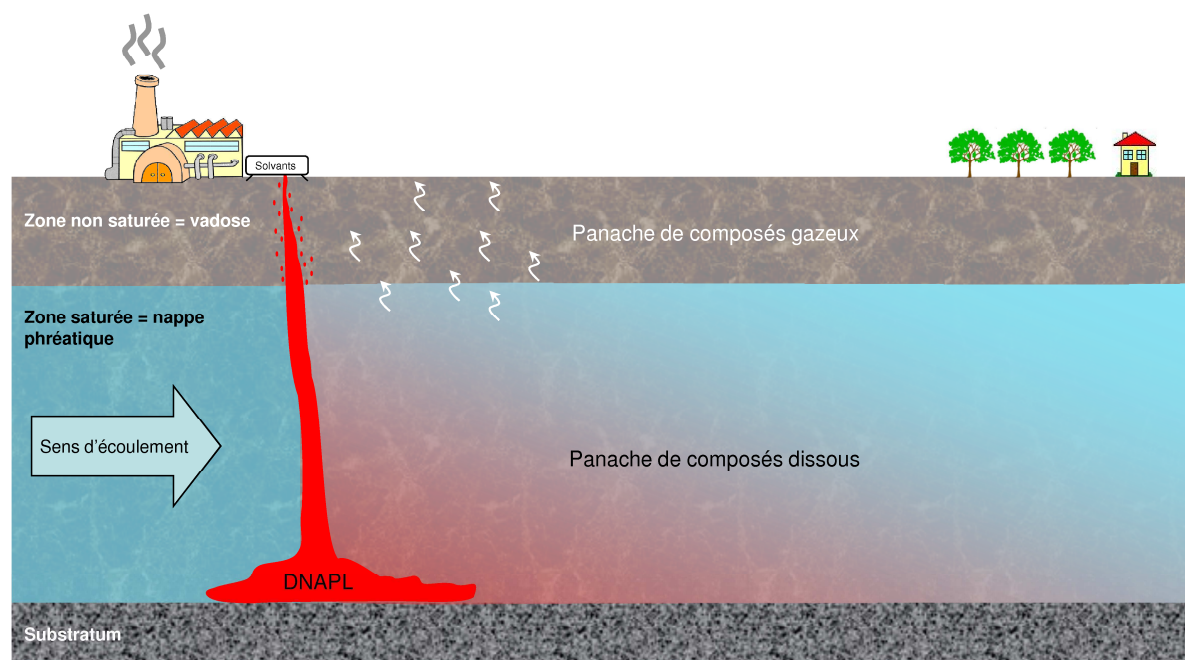


Figure 3 : Migration des chloroéthènes dans le milieu souterrain suite à une contamination d'origine industrielle. DNAPL : Dense Non Aqueous Phase Liquid.

1.3. Propriétés des chloroéthènes et devenir dans l'environnement

1.3.1. Propriétés

Les chloroéthènes sont des composés volatils et appartiennent donc au groupe des Composés Organo Halogénés Volatils (COHV). Leur niveau de volatilité est fonction du nombre de substitutions comme le traduit leur pression de vapeur qui diminue avec celui-ci (Tableau 2). Ainsi, le VC, avec un seul atome de Cl, est le composé le plus volatil, et se trouve même à l'état gazeux à température ambiante, alors que le PCE avec 4 atomes de Cl est le moins volatil. Après libération dans l'environnement, les alcènes chlorés auront donc tendance à se propager dans l'atmosphère (85% du PCE libéré, plus de 99% du DCE). Cependant, lors de contaminations massives (défaut de stockage par exemple), d'autres propriétés physiques vont leur permettre de gagner rapidement les nappes d'eaux souterraines au fond desquelles ils vont s'accumuler et persister (Figure 3) à savoir :

- **La mobilité dans le sol :** les alcènes chlorés sont très mobiles dans le sol et ne s'adsorbent que de manière négligeable sur les particules qui le composent. Ceci se traduit par un coefficient d'adsorption (K_{oc}) relativement faible (Tableau 2). Cette caractéristique explique qu'une fois présents dans le sol, ces composés vont rapidement gagner la nappe d'eau souterraine.
- **La solubilité dans l'eau :** de manière générale, leur solubilité est très faible et diminue avec le nombre de substitutions (Tableau 2). En fonction du niveau de pollution, les nappes phréatiques pourront donc présenter deux phases : une phase aqueuse contenant de faibles quantités de polluants (concentrations inférieures au seuil de solubilité) et, lors de forte pollution, une phase organique supplémentaire dite « huileuse » également appelée « coulant » qui contiendra du produit pur et qui stagnera au niveau du substratum de l'aquifère.
- **La densité :** à l'exception du VC, les alcènes chlorés sont plus denses que l'eau (Tableau 2) et migrent donc par gravité jusqu'au substratum de l'aquifère lors d'une contamination importante de la nappe phréatique.

1.3.2. Devenir des solvants chlorés

L'ensemble des propriétés décrites précédemment font que, suite au déversement sur le sol de solvants chlorés, deux processus distincts et successifs vont se mettre en place

(Figure 3) (données adaptées du guide méthodologique « Atténuation naturelle des composés organo-chlorés aliphatiques dans les aquifères » rédigé dans le cadre du programme ANR MACAOH (Modélisation, Atténuation, Caractérisation dans les Aquifères des Organo-Halogénées, 2001-2006)) :

- **La création d'une zone source de contamination :** le sol (zone non saturée en eau) est un milieu poreux où les phénomènes capillaires jouent un rôle prépondérant. La phase organique va se déplacer par gravité dans cette matrice et gagner l'aquifère, laissant derrière elle une zone à saturation résiduelle (corps d'imprégnation) qui comprend des lentilles de polluants purs fortement dispersées et de faible dimension, mais aussi éventuellement des lentilles de plus grande dimension à l'interface entre des niveaux présentant des contrastes élevés de perméabilité. Au niveau de l'aquifère (zone saturée en eau), la phase organique poursuit son déplacement, toujours par gravité, jusqu'au fond de la nappe où elle s'accumule sous forme de poches de produits purs appelées phase liquide dense non aqueuse ou Dense Non Aqueous Phase Liquid (DNAPLs) pour constituer la zone source de contamination.
- **Le développement d'un panache de contamination :** le (ou les) composé(s) présent(s) dans la zone source de contamination va (vont) se solubiliser dans l'eau de la nappe et entraîner la formation d'un panache de composés dissous mais il(s) peut (peuvent) également se volatiliser dans l'air de la zone non saturée et conduire au développement d'un panache de composés gazeux. Du fait de la lenteur des transferts par dissolution et volatilisation, la disparition complète de la zone source est un mécanisme très long (plusieurs années voire plusieurs dizaines d'années). Sa durée va dépendre non seulement de l'ampleur de la contamination (quantité totale de solvants déversés) mais également de la vitesse d'écoulement de la nappe phréatique. Le panache de composés dissous, en général d'extension largement plus grande que la dimension de la zone source, se développe à partir de la nappe essentiellement par convection. Le panache de composés gazeux, quant à lui, se déplace principalement à la verticale de la zone source et est essentiellement contrôlé par la diffusion. Cependant, un transfert progressif des composés dissous présents dans la nappe vers la zone non saturée puis vers l'atmosphère peut s'effectuer par volatilisation et inversement (en particulier par lessivage des gaz du sol en cas de pluie).

Après contamination de la nappe phréatique, la persistance des chloroéthènes sera extrêmement longue si aucun processus biologique ne se met en place. En revanche, si les conditions biologiques *in situ* sont propices, leur biodégradation en condition anaérobie s'effectuera assez rapidement comme le montre le Tableau 3 qui indique des temps de demi-

Tableau 3 : Temps de demi-vie des différents chloroéthènes observés en fonction des processus de dégradation et concentrations seuils fixées par les autorités sanitaires (adapté de Barbee *et al.*, 1994).

	Demi-vies des chloroéthènes (en années)		Concentrations seuils tolérées dans les eaux de consommation	
	Processus abiotiques	Biodégradation anaérobie	Directive de l'Union Européenne relative à la qualité des eaux destinées à la consommation humaine (1998)	Directives de qualité pour l'eau de boisson de l'OMS (récapitulées dans le "Guidelines for drinking- water quality", 2011)
PCE	$3,8 \times 10^8$ - $9,9 \times 10^8$	34 - 230	10 $\mu\text{g/L}$	40 $\mu\text{g/L}$
TCE	$4,9 \times 10^5$ - $1,3 \times 10^6$	33 - 230	10 $\mu\text{g/L}$	20 $\mu\text{g/L}$
1,1-DCE	$4,7 \times 10^7$ - $1,2 \times 10^8$	81 - 173	non concerné	30 $\mu\text{g/L}$
1,2-DCE	$8,5 \times 10^9$ - $2,1 \times 10^{10}$	88 - 339	non concerné	50 $\mu\text{g/L}$
VC	>10	>60	0,5 $\mu\text{g/L}$	0,3 $\mu\text{g/L}$

vie (Barbee, 1994). Ces derniers sont de l'ordre de quelques dizaines à quelques centaines de jours d'après des calculs faits lors d'expériences réalisées dans des conditions proches de celles d'un aquifère contaminé.

1.4. Risques sanitaires

La quasi-totalité des solvants présente des risques pour la santé. Une exposition à ces molécules, et plus particulièrement à celles présentant des pouvoirs cancérogènes et mutagènes, doit donc être limitée au maximum. Un grand nombre de pathologies d'origine professionnelle a été rapporté suite à une manipulation ou à une exposition prolongée des travailleurs aux solvants chlorés. Ces pathologies peuvent également toucher les particuliers dont les habitations sont situées à proximité de sites industriels contaminés. Dans ce dernier cas, la contamination provient surtout de l'inhalation de vapeurs provenant du panache gazeux ou de l'ingestion d'eau prélevée à partir de captages touchés par un panache de composés dissous. Actuellement, seul le VC est officiellement reconnu comme cancérogène pour l'Homme, les autres chloroéthènes étant classés comme « carcinogènes probables » par le Centre International de Recherche sur le Cancer (CIRC, <http://www.iarc.fr/>) et l'Organisation Mondiale de la Santé (OMS, <http://www.who.int/>). C'est en effet en 1995 que leur toxicité a été réévaluée par le CIRC pour passer du groupe 3 (inclassable quant à sa cancérogénicité pour l'Homme) au groupe 2A (probablement cancérogène). Cependant, leurs effets sur la santé sont bien réels et peuvent toucher, en fonction du composé considéré, le système nerveux central, le foie, le rein, le pancréas, le système circulatoire et le système reproducteur (Ruder, 2006). Ces effets sont essentiellement attribués à l'effet cytotoxique des espèces réactives produites au cours de la dégradation des chloroéthènes par les enzymes présentes dans ces différents tissus (cytochromes P450, GST...) (Lash and Parker, 2001; Lock and Reed, 2006). C'est pourquoi les autorités sanitaires internationales (de l'Union Européenne et de l'OMS) ont fixé des teneurs maximales à ne pas dépasser dans les eaux destinées à la consommation (Tableau 3).

Contrairement à la majorité des molécules chlorées, les chloroéthènes ne sont pas considérés comme des composés fortement bioaccumulables dans les tissus adipeux des animaux qu'ils soient terrestres ou aquatiques. Par conséquent, malgré le manque de données toxicologiques les concernant, le risque d'accumulation de ces molécules le long des chaînes trophiques apparaît relativement faible (<http://www.ineris.fr> ; (Andersen *et al.*, 2008)).

Tableau 4 : Processus microbiens permettant la biodégradation des chloroéthènes.

Composé	Aérobic		Anaérobic	
	Co-métabolisme	Assimilation	Déchloration Oxydative	Déchloration réductrice
PCE				X
TCE	X			X
DCE	X	X	X	X
VC	X	X	X	X

2. Biodégradation des chloroéthènes

Depuis plusieurs décennies, l'accumulation de produits chimiques toxiques dans l'environnement ne cesse d'augmenter. Elle touche toutes les sociétés industrialisées et se propage également, notamment par voie atmosphérique, vers les zones reculées de la biosphère. Les nuisances et les risques pérennes pour les individus ainsi que pour les écosystèmes sont donc extrêmement préoccupants. Heureusement, les microorganismes, et plus particulièrement les procaryotes, grâce à leurs nombreuses machineries enzymatiques et à leurs fortes capacités d'adaptation, sont capables de catalyser efficacement la dégradation d'un très grand nombre de composés organiques (Guerrero and Berlanga, 2006). Ainsi, certaines souches bactériennes ont su évoluer et adapter leurs voies métaboliques pour utiliser de nouveaux substrats tels que les polluants (van der Meer *et al.*, 1992; Janssen *et al.*, 2005). Elles peuvent alors les utiliser comme sources de carbone, comme donneurs ou accepteurs d'électrons dans des processus respiratoires mais également de manière fortuite grâce à l'activité non-spécifique de certaines de leurs enzymes (co-métabolisme) (Mattes *et al.*, 2010; Pant and Pant, 2010). Actuellement, plusieurs bactéries capables de réaliser la dégradation des chloroéthènes ont été identifiées. Elles utilisent des processus très variés qui peuvent aboutir à une dégradation complète ou incomplète de ces molécules.

Quatre voies métaboliques principales de biodégradation des chloroéthènes par les bactéries sont actuellement décrites. Parmi elles, deux nécessitent la présence d'oxygène : le co-métabolisme aérobie et l'assimilation aérobie. *A contrario*, les deux autres, à savoir la déchloration réductrice anaérobie (métabolique ou co-métabolique) et la déchloration oxydative anaérobie, se déroulent en anaérobiose (Mattes *et al.*, 2010). Certaines voies ne concernent qu'un nombre limité de molécules (Tableau 4). Ainsi, le PCE ne peut être dégradé qu'en condition anaérobie par la voie de déchloration réductrice, alors que la transformation des autres chloroéthènes s'effectue indifféremment en présence ou en absence d'oxygène (Tableau 4).

2.1. Le co-métabolisme aérobie

Le co-métabolisme est un processus faisant intervenir plusieurs catégories d'enzymes qui, en plus de leur fonction métabolique classique, vont pouvoir transformer le composé chloré grâce à leur large spectre d'activités. Cependant, au cours de celui-ci, le microorganisme ne tirera aucun bénéfice de son utilisation (Alvarez-Cohen and Speitel, 2001;

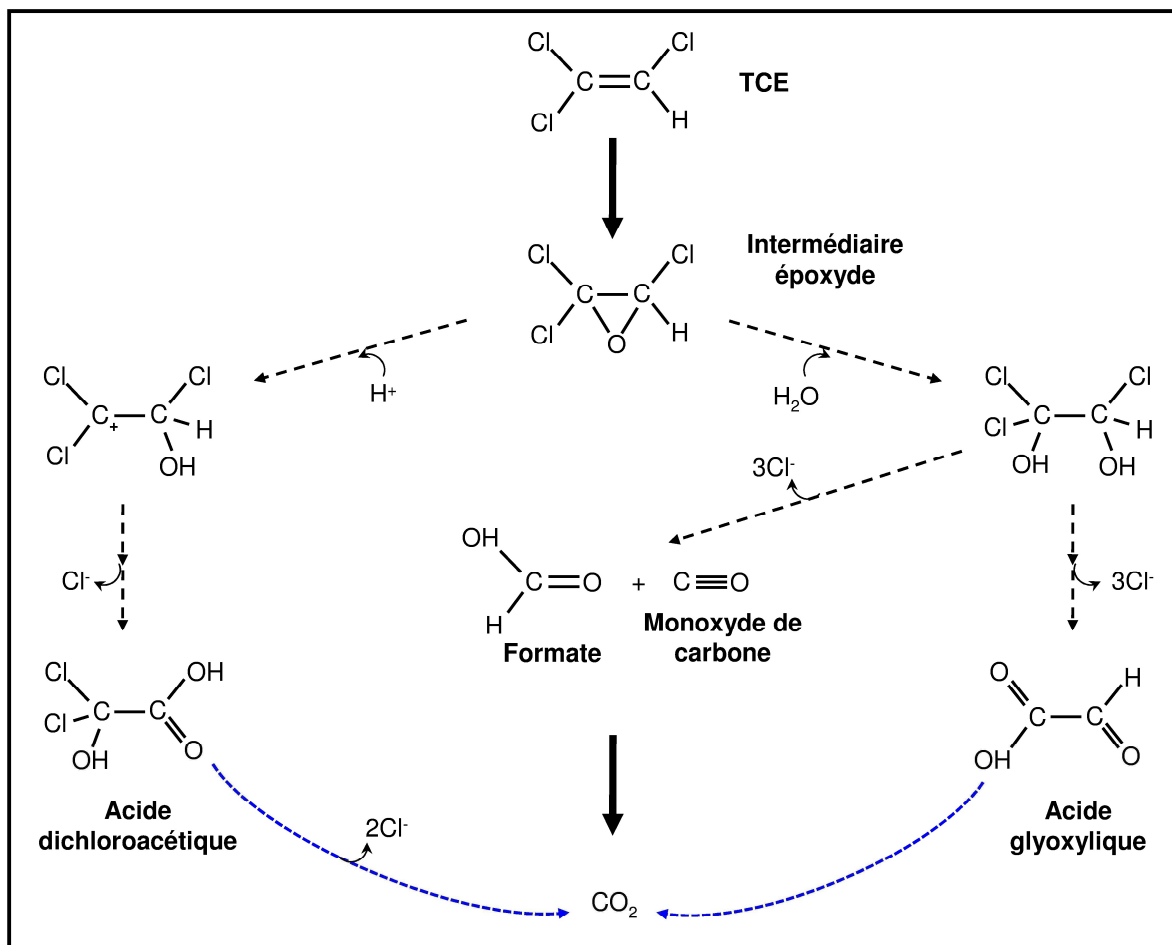


Figure 4 : Voie de biodégradation du trichloroéthylène (TCE) par co-métabolisme aérobie chez les bactéries méthanotrophes (d'après Little *et al.*, 1988). Les flèches pleines indiquent les étapes catalysées par la méthane monooxygénase (MMO), les flèches en pointillées celles qui ont lieu spontanément et les flèches bleues celles qui peuvent être réalisées par d'autres microorganismes hétérotrophes.

Arp *et al.*, 2001; Pant and Pant, 2010). Seuls trois chloroéthènes sur les quatre existants sont dégradés par co-métabolisme aérobie (Tableau 4). Par ailleurs, c'est l'unique voie possible pour la dégradation du TCE en présence d'oxygène.

Ce processus, initialement montré dans des colonnes de sol par Wilson et Wilson (1985), a ensuite été caractérisé chez des bactéries méthanotrophes aussi bien de type I, comme la souche 46-1 (Little *et al.*, 1988), que de type II, comme *Methylosinus trichosporium* OB3b (Oldenhuis *et al.*, 1989; Tsien *et al.*, 1989; Lontoh and Semrau, 1998), *Methylosinus* sp. LAC (Alvarez-Cohen *et al.*, 1992) et *Methylomonas methanica* 68-1 (Koh *et al.*, 1993). Ainsi, les premières enzymes identifiées comme étant capables de dégrader le TCE par co-métabolisme ont comme fonction réelle l'oxydation du méthane : il s'agit des méthane monooxygénases (MMO) solubles (sMMO) et particulaires (pMMO). Au cours de la réaction, le TCE est tout d'abord transformé en époxyde sous l'action de la MMO. Cet intermédiaire réactionnel très instable est ensuite spontanément dégradé en acides dichloroacétique et/ou glyoxylique, voire en composés plus simples comme le formate ou le monoxyde de carbone. Les atomes de chlore sont libérés sous forme d'ions chlorures (Figure 4). Les différents métabolites rejoignent ensuite les voies classiques de biodégradation de la matière organique et sont minéralisés sous forme de CO₂ et d'H₂O. Bien que l'essentiel de la littérature décrive ce processus pour la dégradation du TCE, des réactions identiques sont à l'origine de l'élimination du DCE et du VC (Oldenhuis *et al.*, 1989; Tsien *et al.*, 1989; Pant and Pant, 2010).

Les autres enzymes capables de catalyser l'oxydation co-métabolique des chloroéthènes appartiennent principalement aux groupes des mono- et dioxygénases (Arp *et al.*, 2001). Ainsi, la toluène dioxygénase (TDO) décrite chez *Pseudomonas putida* F1 (Wackett and Gibson, 1988) et les toluène monooxygénases (TMO) décrites chez *Burkholderia cepacia* G4 (Landa *et al.*, 1994), *Pseudomonas mendocina* KR1 (Winter *et al.*, 1989), *P. stutzeri* OX1 (Shim *et al.*, 2001), et *Ralstonia pickettii* PKO1 (Leahy *et al.*, 1996), peuvent dégrader le TCE alors qu'elles ont normalement comme fonction, l'oxydation de composés aromatiques tels que le toluène et le benzène. Les alcènes monooxygénases de *Xanthobacter* sp. Py2 codée (Ensign *et al.*, 1992) et de *Rhodococcus corallinus* B-276 (Saeki *et al.*, 1999), l'ammoniaque monooxygénase de *Nitrosomonas europaea* (Arciero *et al.*, 1989), la butane monooxygénase de *Pseudomonas butanovora* (Hamamura *et al.*, 1997), la propane monooxygénase de *Mycobacterium vaccae* JOB5 (Wackett *et al.*, 1989) et la phénol hydroxylase de *Ralstonia eutropha* JMP134 (Harker and Kim, 1990) possèdent également cette capacité.

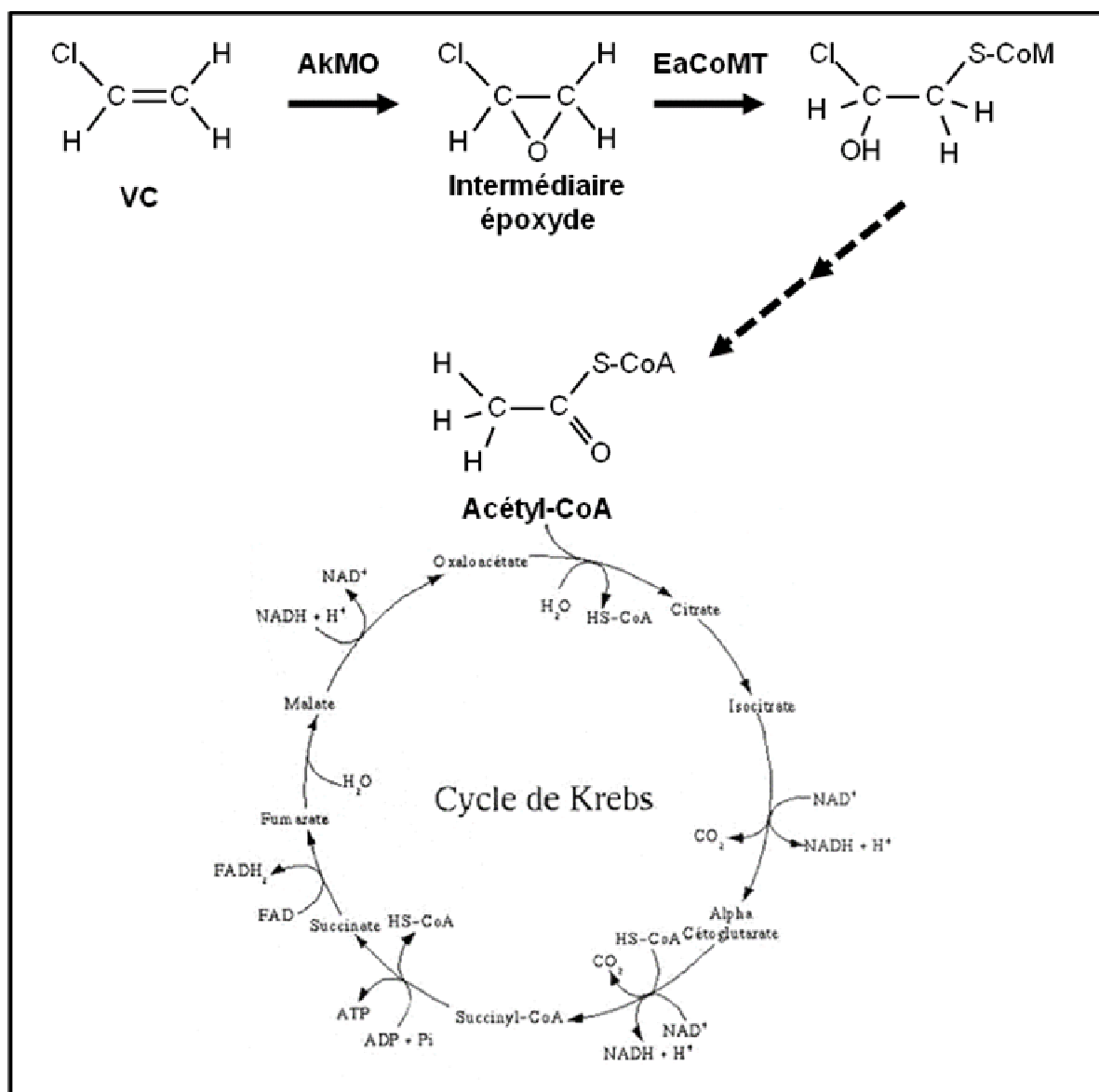


Figure 5 : Mécanisme d'assimilation aérobie du chlorure de vinyle (VC) empruntant la voie classique de dégradation de l'éthylène (d'après Mattes *et al.*, 2010). AkMO : acétylène monooxygénase. EaCoMT : époxyalcane coenzyme M transférase.

Puisqu'il n'est réalisé que de manière fortuite par des enzymes dont ce n'est pas la fonction naturelle, le co-métabolisme est un processus relativement lent comparé aux autres processus métaboliques nécessaires à la croissance des microorganismes. De plus, dans l'environnement, son utilisation pour la biodégradation des chloroéthènes ne peut se faire qu'en présence du substrat naturel de l'enzyme (Alvarez-Cohen and Speitel, 2001). Sa vitesse est également dépendante du nombre de substitutions de la molécule. En effet, les molécules les moins chlorées seront dégradées plus vite et en plus grande quantité (Alvarez-Cohen and Speitel, 2001). Par ailleurs, une concentration trop élevée en chloroéthènes aura pour effet d'inhiber l'activité des oxygénases et pourra même être toxique pour le microorganisme (Broholm *et al.*, 1990; Heald and Jenkins, 1994; Hyman *et al.*, 1995; Ely *et al.*, 1997; Mars *et al.*, 1998). Enfin, d'autres paramètres abiotiques tels que la température, l'humidité et la salinité sont également susceptibles d'influencer l'efficacité de la réaction (Fan and Scow, 1993; Lee *et al.*, 2005; Lee and Liu, 2006).

2.2. L'assimilation aérobie

L'assimilation aérobie ne concerne que le DCE et le VC (Tableau 4). Ce processus métabolique très efficace permet aux microorganismes qui le réalisent d'intégrer ces molécules dans leur métabolisme pour conduire soit à leur utilisation comme source de carbone pour la synthèse de molécules plus complexes (anabolisme), soit à leur utilisation comme source d'énergie en les dégradant jusqu'en CO₂ (catabolisme) (Mattes *et al.*, 2010). Pour cela, ils utilisent la voie classique de dégradation de l'éthylène (Figure 5). Le mécanisme réactionnel n'est pas encore complètement élucidé mais il fait intervenir une alcène monooxygénase (AkMO) qui transforme tout d'abord le composé chloré en son époxyde. Celui-ci est ensuite conjugué au coenzyme M par une époxyalcane coenzyme M transférase (EaCoMT) (van Hylckama Vlieg and Janssen, 2001; Coleman and Spain, 2003a). Le conjugué ainsi formé rejoint ensuite le métabolisme central (cycle de Krebs). Ce processus a été mis en évidence pour la première fois chez la souche *Mycobacterium aurum* L1 (Hartmans *et al.*, 1985; Hartmans and De Bont, 1992) puis ensuite chez de nombreuses souches affiliées à ce genre (Coleman and Spain, 2003b). Depuis, des microorganismes appartenant à d'autres genres et capables de réaliser ce mécanisme ont été décrits tels que *Nocardioides* sp. JS614 (Coleman *et al.*, 2002a; Mattes *et al.*, 2005), *Pseudomonas aeruginosa* MF1 (Verge *et al.*, 2000) et DL1 (Verge *et al.*, 2001), *P. putida* AJ et *Ochrobactrum* sp. TD (Danko *et al.*, 2004), et plus récemment *Ralstonia* sp. TRW1 (Elango *et al.*, 2006).

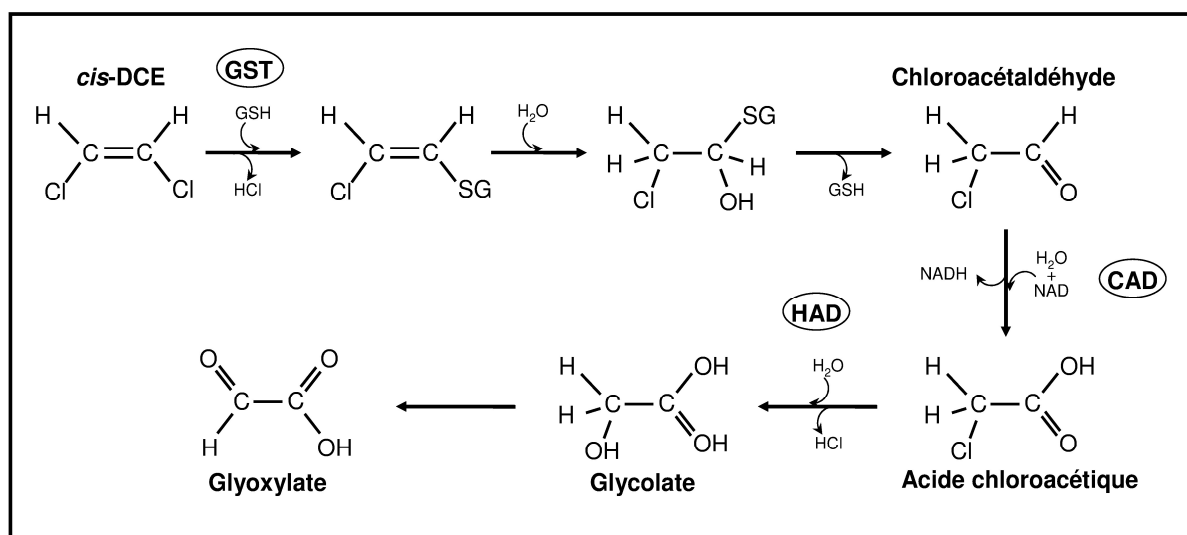


Figure 6 : Mécanisme supposé d'assimilation aérobie du dichloroéthylène (DCE) par la souche *Polaromonas chloroethenica* JS666 (adaptée de Jennings *et al.*, 2009). GSH : Glutathion. GST : glutathion-S-transférase. CAD : chloroacétaldéhyde deshydrogénase. HAD : haloacide déhalogénase.

Pendant de nombreuses années, ce processus n'était décrit que pour la biodégradation du VC. En effet, si les enzymes décrites précédemment étaient également capables d'utiliser le DCE, ceci ne se faisait qu'au cours de processus co-métaboliques (Verge *et al.*, 2002). C'est en 2002, que Coleman *et al.* ont isolé le seul microorganisme connu à ce jour capable d'assimiler le DCE en présence d'oxygène : *Polaromonas chloroethenica* JS666 (Coleman *et al.*, 2002b). Cependant, le séquençage et l'annotation de son génome (Mattes *et al.*, 2008) n'ont pas permis d'identifier de séquences codant pour des protéines homologues aux AkMO et EaCoMT. D'après les travaux réalisés par Jennings *et al.* (2009), le mécanisme enzymatique à l'origine du processus d'assimilation aérobie du DCE serait légèrement différent de celui décrit chez les organismes oxydant le VC. En effet, l'analyse du transcriptome et du protéome de cette souche cultivée avec du *cis*-DCE comme seule source de carbone suggère que la glutathion-S-transférase (GST) catalyserait le clivage de la première liaison C-Cl (déhalogénéation). Cette réaction serait alors suivie d'une hydroxylation du second groupement Cl par une haloacide déhalogénase (HAD) plutôt que d'une époxydation par une monooxygénase (Figure 6).

Parmi les nombreux microorganismes capables d'assimiler le DCE ou le VC en présence d'oxygène, les espèces affiliées au genre *Mycobacterium* sont les plus largement distribuées dans l'environnement (Coleman *et al.*, 2002a). Ces microorganismes semblent particulièrement bien acclimatés aux environnements contaminés par ces polluants et possèdent probablement des capacités d'adaptation qui leur permettent de coloniser plus facilement ces écosystèmes. En effet, il a pu être montré que plusieurs souches de *Mycobacterium* initialement capables d'assimiler de l'éthylène mais pas le VC comme substrat de croissance, acquerraient cette capacité après 2 à 16 mois d'exposition à ce polluant pour finalement l'utiliser comme seule source de carbone et d'énergie (Jin and Mattes, 2008). De plus, ces microorganismes supportent de très faibles concentrations en oxygène, de l'ordre de 0,02 à 0,1 mg/L (Gossett, 2010), ce qui pourrait leur permettre de se développer dans des zones souterraines presque dépourvues d'oxygène et propices à l'accumulation des chloroéthènes (voir paragraphe 1.3).

2.3. Déchloration oxydative anaérobie

Diverses études ont montré que le *cis*-1,2-DCE et le VC pouvaient être convertis de façon oxydative en absence d'oxygène, servant alors de donneurs d'électrons (Bradley and Chapelle, 1996; Bradley and Chapelle, 1998b; Bradley *et al.*, 1998a, b; Bradley *et al.*, 1998c).

Ce mécanisme, appelé déchloration oxydative anaérobie, n'a été observé qu'au sein de milieux réducteurs en présence de divers accepteurs d'électrons (Fe(III), acides humiques, Mn(IV), SO_4^{2-}) seuls ou en mélange (Bradley *et al.*, 2005). Des réactions identiques ont également été décrites dans des milieux méthanogènes (Bradley and Chapelle, 1998a, 1999, 2000). Au cours de ce processus, le *cis*-1,2-DCE et le VC sont complètement minéralisés en CO_2 et en ions chlorure (Cl^-). Cependant, les connaissances concernant les réactions mises en jeu et les microorganismes impliqués sont encore très limitées. Certains travaux suggèrent même que ces observations seraient, en réalité, la conséquence de l'action de bactéries microaérophiles impliquées dans la voie d'assimilation aérobie (paragraphe 2.2) (Gossett, 2010).

Actuellement, les seuls microorganismes qui ont réellement été montrés comme étant capables d'oxyder des chloroéthènes en anaérobiose sont affiliés au genre *Clostridium*. Il s'agit de deux souches, DC1 et KYT-1, isolées de sédiments d'un bassin contenant des lixiviats de décharges provenant de Corée du Sud (Hata *et al.*, 2003; Hata *et al.*, 2004; Kim *et al.*, 2006). Comparée à la souche DC1 qui dégrade le *cis*-1,2-DCE et le VC, la souche KYT-1 montre également des capacités importantes de dégradation des autres isomères du DCE (*trans*-1,2-DCE et 1,1-DCE) mais également du TCE, du PCE et de plusieurs chloroéthanes. Dans ces cultures, l'oxydation du *cis*-1,2-DCE et du VC ne produit pas d'éthylène ni d'éthane mais nécessite l'ajout de glucose, ce qui indique que ce processus n'est pas lié à la croissance du microorganisme (Hata *et al.*, 2003). A ce jour, les voies métaboliques et les enzymes utilisées au cours de ce processus oxydatif ainsi que les produits terminaux de dégradation ne sont toujours pas décrits.

Bien qu'il existe un manque évident de données sur les mécanismes d'oxydation anaérobie des chloroéthènes, la biostimulation de cette voie de dégradation ou la bioaugmentation de certains sites avec des souches ou *consortia* capables de réaliser la déchloration oxydative anaérobie pourrait constituer une stratégie de bioremédiation efficace (Bradley *et al.*, 1998b; Hata *et al.*, 2003).

2.4. La déchloration réductrice anaérobie

La déchloration réductrice est une réaction d'oxydo-réduction qui se déroule en absence d'oxygène et au cours de laquelle le composé chloré perd un groupement halogène et accepte deux électrons (Pemont, 2005; Hiraishi, 2008). Dans le cas des chloroéthènes, cette voie métabolique peut conduire à la déchloration successive du PCE jusqu'en éthylène en

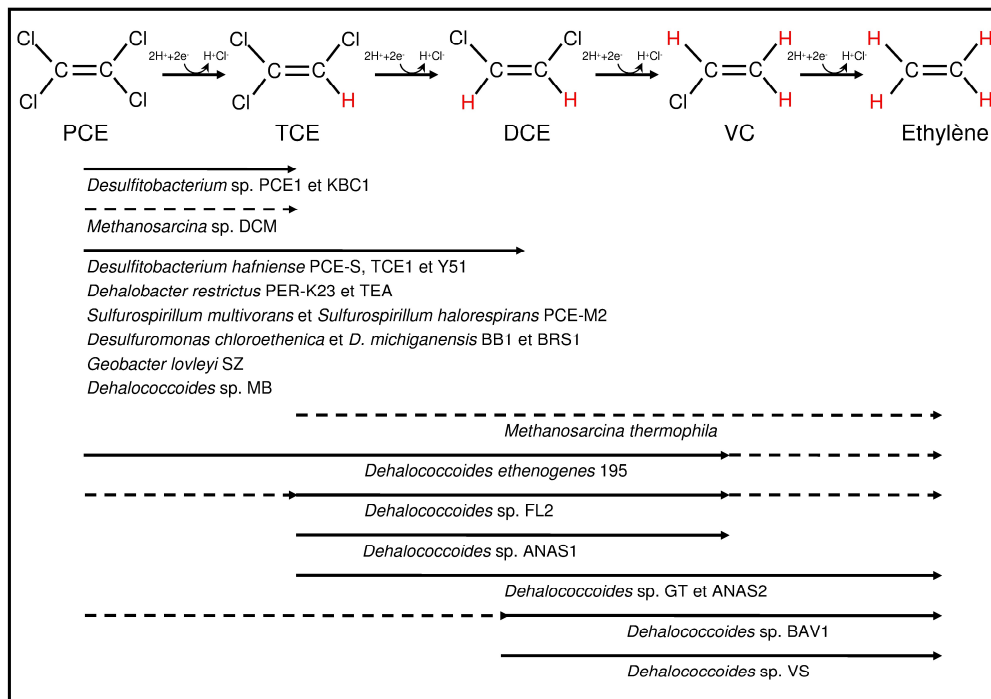


Figure 7 : Déchloration réductrice anaérobie des chloroéthènes par différentes souches procaryotiques. Les flèches pleines indiquent les procédés de déchloration réductrice métabolique (déhalorespiration) et celles en pointillées, ceux de déchloration réductrice co-métabolique (adaptée de Futugami *et al.*, 2008). La souche *Dehalococcoides* sp. MB dégrade le PCE et le TCE en *trans*-1,2-DCE alors que les autres souches réalisant cette étape conduisent à la formation de *cis*-1,2-DCE.

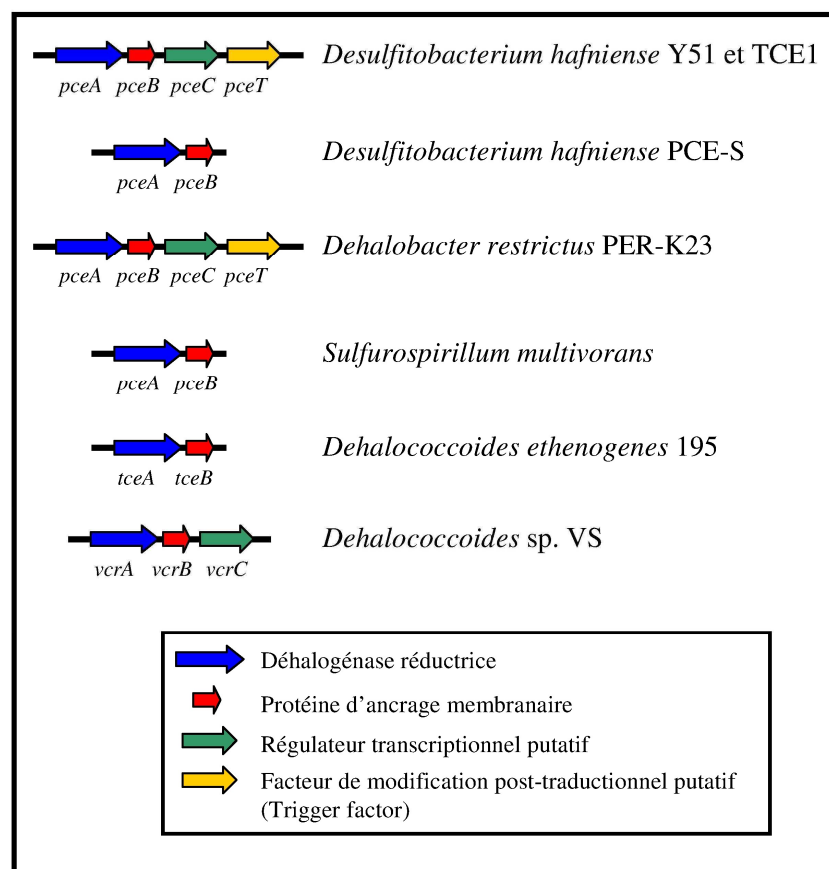


Figure 8 : Organisation des opérons contenant les gènes codant pour les déhalogénases réductrices chez différentes souches déhalorespirantes (d'après Futagami *et al.*, 2008).

passant par les intermédiaires TCE, DCE et VC (Figure 7) (Vogel and McCarty, 1985). Au cours de cette réaction, si l'utilisation du composé chloré est couplée à la production d'énergie par le microorganisme, la déchloration réductrice est dite métabolique et on parlera alors de déhalorespiration (Holliger *et al.*, 1998a). En revanche, si cette réaction est catalysée par des enzymes réalisant d'autres fonctions métaboliques et que le microorganisme n'en tire aucun bénéfice (comme ceci a déjà été décrit dans le paragraphe 2.1), on parlera de déchloration réductrice co-métabolique (Vandecasteele, 2005; Furukawa, 2006; Hiraishi, 2008).

2.4.1. Déchloration réductrice métabolique : la déhalorespiration

La déhalorespiration est un processus respiratoire anaérobie au cours duquel le composé halogéné est utilisé comme accepteur final d'électrons (Holliger *et al.*, 1998a). Elle peut être schématisée par la réaction suivante : $\text{RCl} + \text{H}^+ + 2\text{e}^- = \text{RH} + \text{Cl}^-$. Le premier microorganisme isolé en culture pure capable de réaliser cette réaction est la souche *Desulfomonile tiedjei* DCB-1 qui utilise le 3-chlorobenzoate comme accepteur final d'électrons (DeWeerd *et al.*, 1990). Depuis, de nombreuses souches bactériennes déhalorespirantes ont été identifiées (Smidt and de Vos, 2004; Hiraishi, 2008). Elles réalisent, grâce à un mécanisme d'action très similaire, la dégradation de composés halogénés variés comme des alcanes chlorés (Groster and Edwards, 2006), des chloroéthènes (Chen, 2004), des chlorobenzènes (Adrian *et al.*, 2000), des chlorophénols (Adrian *et al.*, 2007), des éthers diphényles polybromés (PBDEs) (He *et al.*, 2006), des polychlorobiphényles (PCB) (Field and Sierra-Alvarez, 2008) ou encore des dibenzo-p-dioxines et furanes polychlorés (PCDD/Fs) (Bunge *et al.*, 2008; Bunge and Lechner, 2009). Nous nous focaliserons, par la suite, uniquement sur les souches capables de dégrader les solvants chlorés de la famille des chloroéthènes.

a) Eléments intervenant dans la déhalorespiration

L'activité catabolique lors de la déhalorespiration est assurée par des enzymes très spécifiques : les déhalogénases réductrices. Ces enzymes sont codées par des gènes organisés en opérons qui contiennent systématiquement un gène codant pour la sous-unité catalytique (*rdhA*) et un gène codant pour une petite protéine d'ancrage membranaire (*rdhB*) dont la nomenclature est étroitement liée au substrat utilisé (*pceAB*, *tceAB*, *vcrAB*...) (Futagami *et al.*, 2008) (Figure 8). Les déhalogénases réductrices sont donc des protéines hétérodimériques

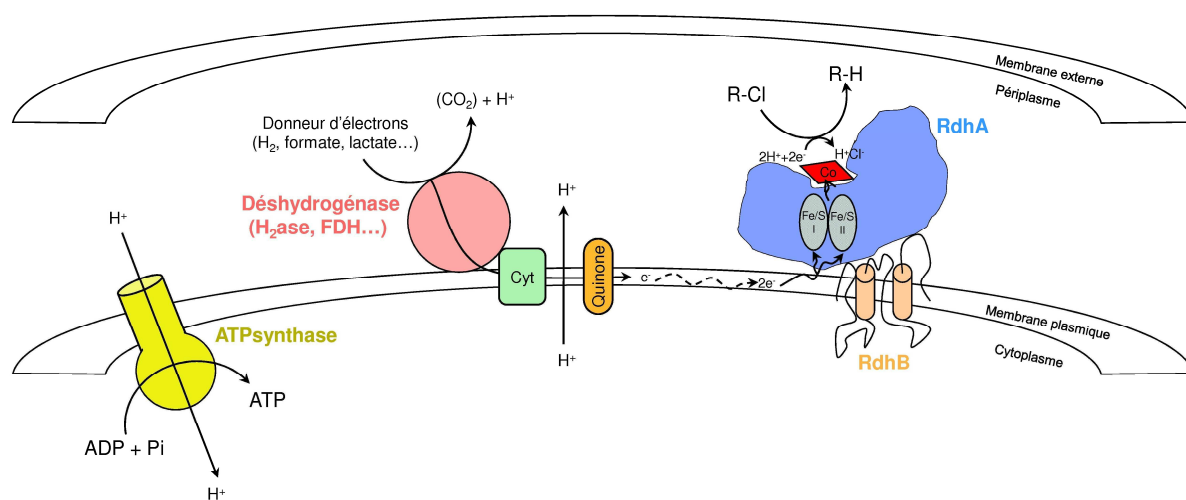


Figure 9 : Mécanisme réactionnel de la déhalorespiration (d'après Holliger et Schumacher, 1994 et Villemur *et al.*, 2006). FDH : formate deshydrogénase. H₂ase : hydrogénase. Cyt : cytochrome. Co : cobalamine.

(sous-unités RdhA et RdhB) généralement associées à la membrane plasmique. Elles peuvent être localisées soit face au périplasma (Suyama *et al.*, 2002; John *et al.*, 2006) soit face au cytoplasme (Schumacher and Holliger, 1996). La plupart des RdhA sont caractérisées par une séquence signal d'adressage membranaire de type Tat (pour Twin-Arginine Translocation), un site de liaison pour un corrinoïde (la vitamine B12, également appelée cobalamine, ou un de ses dérivés) au niveau du centre catalytique et deux motifs permettant de lier des clusters fer-soufre (Smidt and de Vos, 2004; Futagami *et al.*, 2008). Cette architecture va permettre de capter les électrons produits par une déhydrogénase (H₂ase, Formate déshydrogénase en fonction des souches microbiennes) qui sont nécessaires à la réaction d'oxydo-réduction.

Bien que le mécanisme réactionnel n'ait pas encore été finement caractérisé, certains auteurs ont proposés des modèles de déhalorespiration très réalistes (Holliger *et al.*, 1998a; Villemur *et al.*, 2006) (Figure 9). Les électrons seraient transférés à travers la membrane plasmique sous l'action de cytochromes et de quinones jusqu'à la déhalogénase. Au sein de celle-ci, les clusters fer-soufre permettraient leur transfert jusqu'au site catalytique où ils seraient utilisés pour réduire le corrinoïde. Ce dernier, fortement réduit, perdrait ses électrons au contact du composé chloré conduisant à la libération d'un ion Cl⁻ et à son remplacement par un atome d'hydrogène sur la molécule (Fetzner, 1998; Futagami *et al.*, 2008). Le potentiel électrochimique de protons créé par le transfert d'électrons du donneur vers la déhalogénase serait finalement exploité par la bactérie pour synthétiser de l'ATP grâce à l'activité d'une ATP synthase.

b) Diversité des microorganismes déhalorespirants et spécificité de leurs enzymes

A ce jour, plusieurs souches bactériennes déhalorespirantes ont été isolées en culture pure ou en *consortia* à partir d'environnements contaminés par des composés organohalogénés variés. Il s'agit de bactéries aussi bien Gram-positives que Gram-négatives. Elles sont affiliées à trois phyla majeurs, les *Firmicutes*, les *Proteobacteria* et les *Chloroflexi*, sans qu'aucun lien ne puisse être établi entre leur affiliation taxonomique et leurs capacités de dégradation (Smidt and de Vos, 2004; Hiraishi, 2008). Ces souches sont séparées en deux grands types physiologiques selon leurs aptitudes ou non à utiliser des accepteurs d'électrons autres que des molécules halogénées. On distinguera donc les déhalorespirants facultatifs des déhalorespirants obligatoires (Hiraishi, 2008). Les microorganismes déhalorespirants obligatoires sont uniquement affiliés au phylum des *Chloroflexi* (genres *Dehalococcoides*

Tableau 5 : Donneurs et accepteurs d'électrons utilisés par les principaux microorganismes déhalorespirants.

Organisme	Phylum (classe)	Type métabolique	Donneurs d'électrons	Accepteurs d'électrons
<i>Geobacter lovleyi</i> SZ	<i>Proteobacteria</i> (delta)	Facultatif	H ₂ , pyruvate, acétate	Fumarate, nitrate, Fe(III), malate, Mn(IV), U(VI), soufre élémentaire, PCE, TCE
<i>Desulfuroronas chloroethenica</i>			Pyruvate, acétate	Fumarate, Fe(III), polysulfite, nitriloacétate, PCE, TCE
<i>Desulfuromonas michiganensis</i> (souches BB1 et BRS1)			Pyruvate, lactate, acétate, succinate, fumarate, malate	Fumarate, malate, Fe(III), soufre élémentaire, 1,1,2,2-PCA, PCE, TCE
<i>Sulfurospirillum multivorans</i>	<i>Proteobacteria</i> (epsilon)		H ₂ , pyruvate, lactate, formate, éthanol, glycérol	Fumarate, nitrate, selenate, arsenate, PCE, TCE
<i>Sulfurospirillum halorespirans</i> PCE-M2			H ₂ , pyruvate, lactate, formate	Fumarate, nitrate, nitrite, sélénate, arsenate, soufre élémentaire, PCE, TCE
<i>Desulfitobacterium hafniense</i> Y51	<i>Firmicutes</i>		Pyruvate, lactate, formate	Fumarate, nitrate, sulfate, PCE, TCE
<i>Desulfitobacterium</i> sp. TCE1			H ₂ , pyruvate, lactate, formate, butyrate, éthanol, crotonate	Fumarate, nitrate, sulfite, thiosulfate, PCE, TCE
<i>Desulfitobacterium</i> sp. PCE-S			Formate, pyruvate	Fumarate, sulfite, PCE, TCE
<i>Desulfitobacterium</i> sp. PCE1			Pyruvate, lactate, formate, butyrate, succinate	Chlorophénols, fumarate, sulfite, thiosulfate, PCE
<i>Desulfitobacterium</i> sp. KBC1			Pyruvate, lactate, formate, butyrate	Fumarate, sulfite, thiosulfate, PCE
<i>Dehalobacter restrictus</i> (souches PER-K23 et TEA)		Obligatoire	H ₂	PCE, TCE
<i>Dehalococcoides</i> (toutes les souches)	<i>Chloroflexi</i>		H ₂	PCE, TCE, DCE et/ou autres molécules organohalogénées

Tableau 6 : Activités des différentes déhalogénases réductrices identifiées chez des bactéries déhalorespirantes.

Enzymes	Microorganismes	Fonction
PceA	<i>Desulfitobacterium</i> sp. PCE1 <i>Dehalococcoides ethenogenes</i> 195	PCE→TCE
PrdA	<i>Desulfitobacterium</i> sp. KBC1	
PceA	<i>Desulfitobacterium hafniense</i> Y51, TCE1 et PCE-S <i>Dehalobacter restrictus</i> PER-K23 <i>Sulfurospirillum multivorans</i>	PCE→TCE→ <i>cis</i> -DCE
MbrA	<i>Dehalococcoides</i> sp. MB	PCE→TCE→ <i>trans</i> -DCE
TceA	<i>Dehalococcoides ethenogenes</i> 195	TCE→ <i>cis</i> -DCE→VC→Éthylène
VcrA	<i>Dehalococcoides</i> sp. VS	<i>cis</i> -DCE→VC→Éthylène
BvcA	<i>Dehalococcoides</i> sp. BAV1	

(Maymo-Gatell *et al.*, 1997) et *Dehalogenimonas* (Moe *et al.*, 2009)) et à celui des *Firmicutes* (genre *Dehalobacter* (Holliger *et al.*, 1998b)). Ces microorganismes sont très spécialisés puisqu'ils sont capables de tirer leur énergie uniquement de la déchloration réductrice avec comme seuls accepteurs d'électrons des molécules organohalogénées et comme seul donneur d'électrons, l'hydrogène (H₂). Tous les isolats décrits actuellement correspondent à des bactéries mésophiles et neutrophiles vivant dans des habitats d'eau douce et terrestres contaminés par des chloroéthènes (Hiraishi, 2008). En revanche, les souches déhalorespirantes facultatives possèdent des métabolismes variables et diffèrent énormément entre elles dans le nombre d'accepteurs et de donneurs d'électrons qu'elles sont capables d'utiliser (Tableau 5). Les genres décrits jusqu'à aujourd'hui appartiennent aux classes des *Deltaproteobacteria* (*Desulfuromonas* (Krumholz, 1997; Sung *et al.*, 2003) et *Geobacter* (Sung *et al.*, 2006b; Nevin *et al.*, 2007)) et des *Epsilonproteobacteria* (*Sulfurospirillum* (Scholz-Muramatsu *et al.*, 1995; Luijten *et al.*, 2003)), et au phylum des *Firmicutes* (*Desulfitobacterium* (Gerritse *et al.*, 1996; Gerritse *et al.*, 1999; Suyama *et al.*, 2001; Finneran *et al.*, 2002; Tsukagoshi *et al.*, 2006; Villemur *et al.*, 2006)). Quel que soit le type physiologique, chaque souche déhalorespirante possède ses propres capacités métaboliques en fonction de la spécificité de substrat de ses enzymes. En particulier, seules des bactéries du genre *Dehalococcoides* peuvent réaliser la fin de la voie de déchloration réductrice, permettant de métaboliser le DCE et le VC (Figure 7) (Maymo-Gatell *et al.*, 1997; Maymo-Gatell *et al.*, 2001; Cupples *et al.*, 2003; He *et al.*, 2003; He *et al.*, 2005; Sung *et al.*, 2006a; Lee *et al.*, 2011).

Il existe une grande diversité de déhalogénases réductrices dont certaines ont pu être entièrement ou partiellement purifiées et la séquence de leur gène identifiée (Tableau 6). Les enzymes capables de dégrader le PCE sont des PCE déhalogénases. Il en existe plusieurs types et elles peuvent porter des noms différents en fonction des souches bactériennes à partir desquelles elles ont été isolées :

- Les PCE déhalogénases catalysant la dégradation du PCE en TCE. Il s'agit des PceA isolées chez *Desulfitobacterium* sp. PCE1 (van de Pas *et al.*, 2001) et *Dehalococcoides ethenogenes* 195 (Magnuson *et al.*, 1998) ainsi que de la PrdA de *Desulfitobacterium* sp. KBC1 (Tsukagoshi *et al.*, 2006).
- Les PCE déhalogénases catalysant la dégradation du PCE jusqu'en *cis*-DCE via le TCE qui sont toutes appelées PceA. La majorité de ces enzymes présente des séquences protéiques très similaires (96 à 99% d'identité). Elles ont été identifiées chez les souches Y51 (Suyama *et al.*, 2002), TCE1 (van de Pas *et al.*, 2001), PCE-S (Miller *et al.*, 1998) de

Desulfitobacterium hafniense et chez la souche *Dehalobacter restrictus* PER-K23 (Maillard *et al.*, 2003). Seule la PceA de *Sulfurospirillum multivorans* (Neumann *et al.*, 1996) est très éloignée puisqu'elle ne présente que 27% d'identité avec ces dernières.

- La PCE déhalogénase catalysant la dégradation du PCE en *trans*-DCE, appelée MbrA, qui a été récemment identifiée chez la souche *Dehalococcoides* sp. MB (Chow *et al.*, 2010).

Une approche basée sur l'utilisation d'amorces dégénérées ciblant des régions conservées de plusieurs PceA a également permis la mise en évidence de quatre déhalogénases réductrices putatives supplémentaires nommées RdhA chez *Desulfitobacterium* sp. PCE1 et *Sulfurospirillum multivorans* et RdhA1 et RdhA2 chez *Dehalobacter restrictus* PER-K23 (Regeard *et al.*, 2004). De plus, grâce au séquençage et à l'annotation de nombreux génomes bactériens, de nouveaux gènes codant potentiellement pour des PceA ont été découverts. C'est le cas notamment de deux séquences identifiées chez *Shewanella sediminis* HAW-EB3 (numéros d'accension des protéines correspondantes : ABV36709 et ABV36340), de deux autres chez *Geobacter lovleyi* SZ (numéros d'accension : ACD96581 et ACD96583) et d'une chez *Vibrio* sp. RC586 (numéro d'accension : ZP_06078330). Cependant, leurs activités restent encore à confirmer.

Comme les PCE déhalogénases n'interviennent que dans les premières étapes de dégradation du PCE jusqu'en TCE ou DCE, il a été nécessaire de rechercher les enzymes catalysant la fin de voie jusqu'en éthylène, seul composé non toxique. Ces enzymes, appartenant également à la famille des déhalogénases, n'ont été identifiées que chez des souches affiliées au genre *Dehalococcoides*:

- La TCE déhalogénase (TceA) est la seule enzyme catalysant la déchloration réductrice du TCE en DCE puis en VC et peut, par un processus co-métabolique, dégrader le VC en éthylène (Magnuson *et al.*, 1998; Magnuson *et al.*, 2000). Cette protéine, codée par le gène *tceA*, a été caractérisée chez la souche *D. ethenogenes* 195 mais des gènes codant pour des protéines homologues ont également été retrouvés chez les souches FL2 (He *et al.*, 2005) et ANAS1 (Lee *et al.*, 2011) (respectivement 99 et 97% d'identité protéique).
- Les VC réductases utilisent spécifiquement le DCE et le VC comme accepteur final d'électrons pour produire de l'éthylène. La première protéine identifiée capable de faire cette réaction est appelée VcrA et est codée par le gène *vcrA* chez la souche *Dehalococcoides* sp. VS (Muller *et al.*, 2004). Depuis, des gènes codant pour des homologues de cette protéine ont également été retrouvés chez les souches *Dehalococcoides* sp. GT (Sung *et al.*, 2006a) et ANAS2 (Lee *et al.*, 2011) (respectivement 97 et 98% d'identité protéique). Un second gène appelé *bvcA* et identifié par des approches de PCR basées sur l'utilisation d'amorces

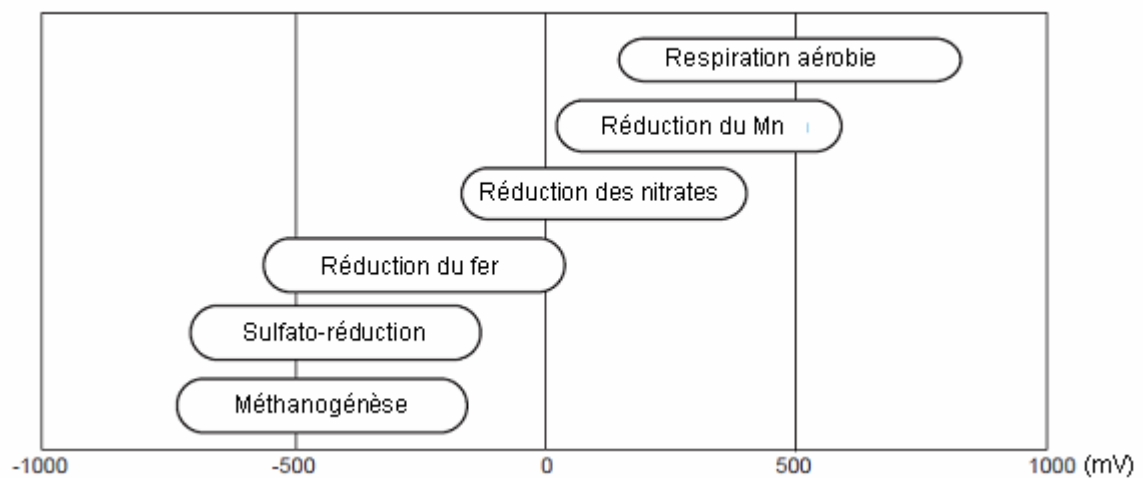


Figure 10 : Potentiels redox caractérisant les principaux processus respiratoires identifiés chez les microorganismes (exprimés en millivolts) (d'après Byl et Williams, 2000). Mn : manganèse.

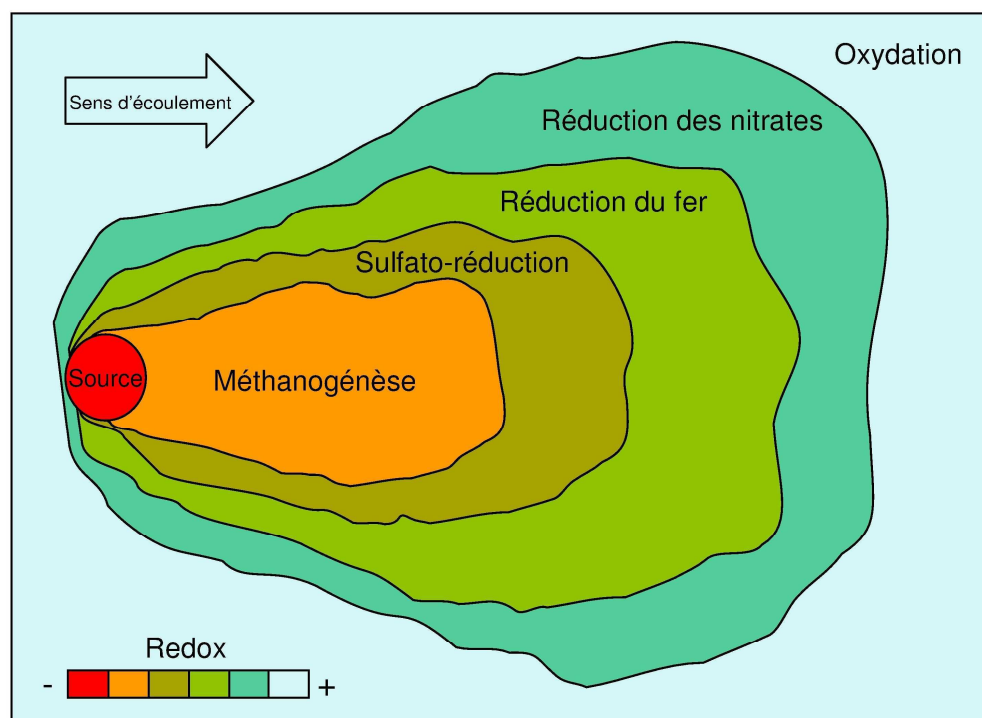


Figure 11 : Succession des différents processus respiratoires le long du panache dans un aquifère pollué par des chloroéthènes (d'après la thèse de F. Nex, 2004).

dégénérées chez la souche *Dehalococcoides* sp. BAV1 pourrait coder pour une autre VC réductase putative : BvcA. Cependant sa séquence protéique est très éloignée de celle de VcrA (39% d'identité) (Krajmalnik-Brown *et al.*, 2004).

c) Conditions physico-chimiques et biologiques propices à la déhalorespiration

La déchloration réductrice est une réaction d'oxydo-réduction. Ainsi, le déclenchement d'un tel processus va être sous le contrôle du potentiel redox du milieu considéré et, en fonction de sa valeur, les populations microbiennes catalysant les réactions les plus énergétiques seront favorisées (Kao *et al.*, 2003). Les principaux accepteurs finaux d'électrons retrouvés sont, par ordre décroissant de pouvoir énergétique : l'oxygène (O₂), le manganèse (MnIV), le nitrate (NO₃⁻), le fer (FeIII), le sulfate (SO₄²⁻) et le dioxyde de carbone (CO₂) (Heimann *et al.*, 2009). Le processus de déchloration réductrice des chloroéthènes est généralement observé pour une gamme de potentiels redox propices à la méthanogénèse et à la sulfato-réduction (Technical protocol for evaluating natural attenuation of chlorinated solvents in groundwater, EPA, 1998 ; (Vogel *et al.*, 1987) (Figure 10).

Par ailleurs, il a été observé que dans un aquifère contaminé par des chloroéthènes, les conditions redox varient le long du panache. Ainsi, l'éloignement de la zone source de contamination s'accompagne généralement d'une diminution du potentiel réducteur (Lovley, 2003) (Figure 11).

La nature du ou des donneurs d'électrons présents dans le milieu va également jouer un rôle déterminant (Azizian *et al.*, 2010). En effet, si la plupart des bactéries déhalorespirantes capables de déchlorer le PCE et le TCE en DCE utilisent une grande variété de donneurs d'électrons issus de la fermentation de la matière organique (lactate, pyruvate, butyrate, acétate, formate, glycérol, éthanol...), la conversion de ce dernier jusqu'en éthylène semble être strictement dépendante de l'H₂ (Lee *et al.*, 2007). Il est donc considéré comme le principal donneur d'électron intervenant dans le processus de déchloration réductrice (Damborsky, 1999; Chen, 2004). Or, comme il joue également un rôle crucial dans plusieurs autres métabolismes anaérobies (dénitrification, sulfato-réduction, réduction du fer, méthanogénèse), la demande pour ce composé est extrêmement forte (Heimann *et al.*, 2009). En témoigne la large distribution des hydrogénases au sein des populations microbiennes, enzymes responsables de son oxydation (Vignais and Billoud, 2007). Les différents acteurs hydrogénotrophes peuvent donc entrer en compétition lorsque l'H₂ devient un élément limitant (Yang and McCarty, 1998). Heureusement, les microorganismes déhalorespirants

Tableau 7 : Concentrations seuils en H₂ nécessaires aux groupes microbiens hydrogénotrophes en fonction de leur métabolisme (d'après Luitjen *et al.*, 2004).

Processus	[H ₂] seuil
Acétogénèse	>350 nM
Méthanogénèse	5-100 nM
Sulfato-réduction	1-10 nM
Ammonification	0,015-0,025 nM
Réduction des nitrates	<0,05 nM
Réduction du manganèse	<0,05 nM
Réduction du fer	0,1-0,8 nM
Dehalorespiration ¹	<0,3 nM
Réduction du PCE et du TCE ²	0,6-0,9 nM
Réduction du <i>cis</i> -DCE ²	0,1-2,5 nM
Réduction du VC ²	2-24 nM

¹ en culture pure ; ² en culture mixtes ou dans l'environnement

présentent pour ce dernier une meilleure affinité que les microorganismes fonctionnant dans les mêmes conditions redox (sulfato-réducteurs, méthanogènes). Ils sont donc avantagés lorsqu'il devient limitant (Tableau 7) (Luijten *et al.*, 2004).

Le bon déroulement de la déhalorespiration nécessite non seulement des bactéries déhalorespirantes mais aussi l'établissement d'interactions syntrophiques entre ces populations et d'autres membres de la communauté microbienne, notamment, *via* des transferts d'H₂. Ainsi, il a été montré, à partir d'une co-culture, que la déchloration du PCE par *Desulfitobacterium hafniense* TCE1 était fortement dépendante de l'activité sulfato-réductrice de *Desulfovibrio fructosivorans* (Drzyzga and Gottschal, 2002). De même, les méthanogènes acétoclastes appartenant au genre *Methanosarcina* ont un effet bénéfique sur la déchloration réductrice du VC en éthylène car elles sont capables de métaboliser l'acétate en CH₄ mais aussi de l'oxyder en CO₂ et en H₂. Ce dernier constitue alors le donneur d'électrons nécessaire aux *Dehalococcoides* pour réaliser la déhalorespiration (Heimann *et al.*, 2006). Enfin, pour les souches de *Dehalococcoides* incapables de synthétiser la vitamine B12, la présence de certaines espèces comme *Acetobacterium woodii* permet de combler ce manque. En effet, ces microorganismes produisent ce co-facteur indispensable au mécanisme de respiration des *Dehalococcoides* et donc à leur développement (He *et al.*, 2007a).

2.4.2. Déchloration réductrice co-métabolique

La déchloration réductrice co-métabolique a été observée chez divers groupes bactériens. Les premiers concernés sont des déhalorespirants obligatoires qui métabolisent, grâce à leurs déhalogénases réductrices, des composés chlorés autres que ceux normalement ciblés (Figure 7). Ainsi, la TceA de *Dehalococcoides ethenogenes* 195, responsable de la dégradation du TCE en VC, est également susceptible de catalyser de manière fortuite la déchloration des molécules de VC en éthylène (Magnuson *et al.*, 2000).

Hormis les organismes déhalorespirants, plusieurs bactéries appartenant à des groupes fonctionnels variés tels que les méthanogènes, les acétogènes, les sulfato-réducteurs et les réducteurs de fer, sont capables de catalyser des réactions de déchloration réductrice co-métaboliques (Holliger and Schraa, 1994; El Fantroussi *et al.*, 1998; Fetzner, 1998). La molécule organohalogénée est alors utilisée comme accepteur d'électrons secondaire, le rôle d'accepteur d'électrons primaire étant assuré par des accepteurs inorganiques comme le CO₂, le sulfate ou le Fe^{III}. La plupart des activités co-métaboliques décrites à ce jour interviennent dans la dégradation d'alcanes halogénés. Cependant, quelques cas de déchloration réductrice

des chloroéthènes par ce processus ont été rapportés. Ainsi, la souche *Methanosarcina* sp. DCM peut dégrader le PCE en TCE en conditions méthanogènes par un mécanisme enzymatique encore non élucidé (Fathepure and Boyd, 1988). L'espèce *M. thermophila*, quant à elle, catalyse la déchloration du TCE jusqu'en éthylène grâce à un complexe enzymatique, appelé carbone monoxyde déshydrogénase (CO Déshydrogénase), qui intervient normalement dans la voie métabolique de fermentation de l'acétate (Jablonski and Ferry, 1992; Maupin-Furlow and Ferry, 1996).

Comparé à la déchloration réductrice métabolique (déhalorespiration), le processus de déchloration réductrice co-métabolique est beaucoup moins efficace et les vitesses de dégradation observées sont nettement inférieures (Fetzner, 1998).

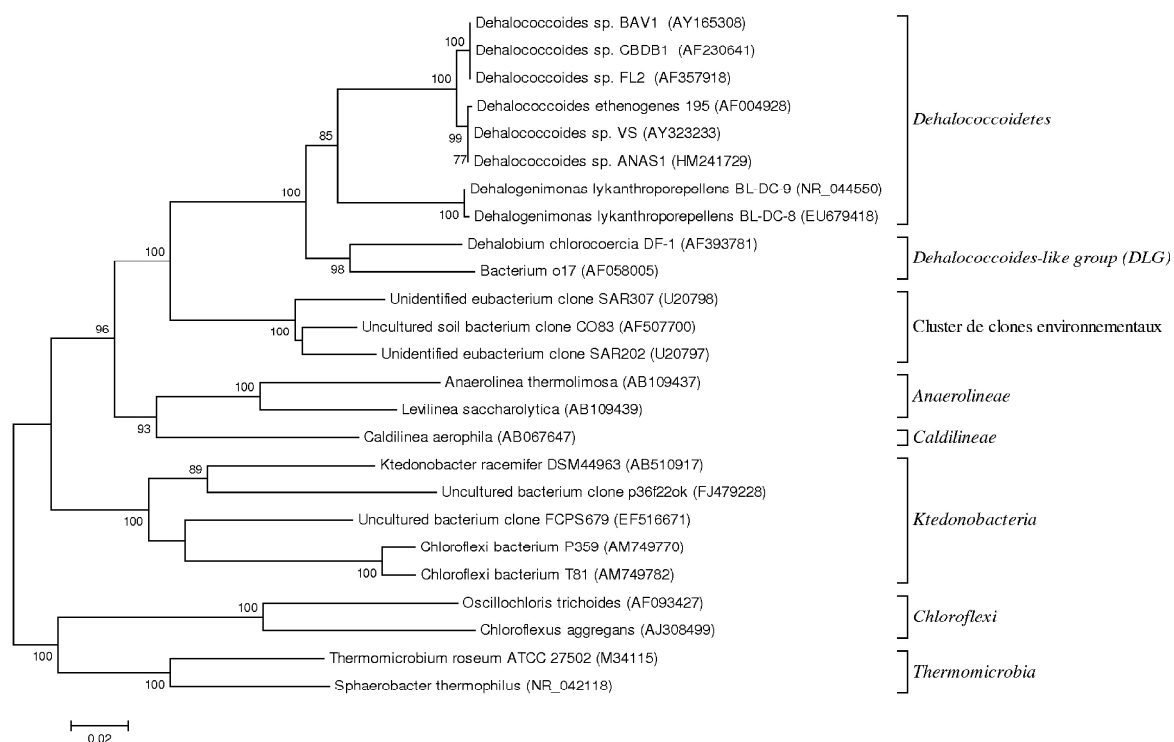


Figure 12 : Arbre phylogénétique représentant les huit groupes du phylum des *Chloroflexi*, soit les sept sous-phyla déjà décrits et le groupe supplémentaire des *Dehalococcoides*-like (DLG). L'arbre a été construit avec le logiciel MEGA 4 (Tamura *et al.*, 2007) sur la base des séquences ARNr 16S dont le numéro d'accèsion est indiqué entre parenthèses, grâce à l'application de la méthode des plus proches voisins (Neighbour Joining) et d'un bootstrap de 100.

3. Les Dehalococcoides

Sur l'ensemble des microorganismes capables de participer à la biodégradation des chloroéthènes, un groupe bactérien suscite depuis de nombreuses années un intérêt tout particulier de la part de la communauté scientifique. Il s'agit des *Dehalococcoides*, seul groupe microbien connu à l'heure actuelle pour dégrader l'ensemble des chloroéthènes par le processus de déhalorespiration (Tas *et al.*, 2010). Les *Dehalococcoides* sont affiliés aux *Chloroflexi*. Ce phylum, anciennement connu sous la dénomination de bactéries vertes non sulfureuses, forme un groupe très hétérogène avec des représentants qui sont largement répandus dans la quasi totalité des écosystèmes (sol, eau douce, écosystèmes marins...etc) et qui présentent une grande diversité de capacités métaboliques (Yamada and Sekiguchi, 2009). En effet, certaines *Chloroflexi* sont capables de réaliser de la photosynthèse anaérobie comme celles appartenant aux genres *Oscillochloris* (Keppen *et al.*, 2000) et *Chloroflexus* (Pierson and Castenholz, 1974) alors que d'autres, comme celles du genre *Thermomicrobium*, sont des organotrophes thermophiles (Jackson *et al.*, 1973).

Ce phylum des *Chloroflexi*, divisé en huit sous-phyla distincts dont sept sont actuellement décrits (Yabe *et al.*, 2010) (Figure 12), renferme le plus grand nombre de microorganismes déhalorespirants (Hiraishi, 2008). Ces derniers comprennent les bactéries des genres *Dehalococcoides* et *Dehalogenimonas*, qui forment actuellement la classe des *Dehalococcoidetes* (Hugenholtz and Stackebrandt, 2004), et un groupe formant un cluster affilié à aucun des sept sous-phyla décrits, les *Dehalococcoides*-like (DLG) (Hiraishi, 2008).

3.1. Caractéristiques des souches affiliées au genre *Dehalococcoides*

Bien que le genre *Dehalococcoides* n'ait pas encore été reconnu officiellement dans la taxonomie des procaryotes (<http://www.bacterio.cict.fr/>), cette nomenclature est couramment utilisée dans la littérature. Une affiliation taxonomique plus fine des membres de ce groupe devrait voir le jour prochainement grâce à l'enrichissement récent de la collection ATCC (<http://www.lgcstandards-atcc.org/>) et la mise à disposition de 3 souches affiliées à celui-ci. Historiquement, la première souche appartenant au genre *Dehalococcoides* a été décrite en 1997 et nommée *Dehalococcoides ethenogenes* 195 (Maymo-Gatell *et al.*, 1997). Encore aujourd'hui, elle est la seule souche décrite comme étant capable de déchlorer totalement le PCE jusqu'en éthylène. Depuis cette première découverte, un grand nombre de souches affiliées à ce groupe ont été décrites dans la littérature comme les souches CBDB1 (Adrian *et*

Tableau 8 : Liste des molécules organohalogénées dégradées par les différentes souches de *Dehalococcoides* et des métabolites qu'elles produisent au cours du processus de déhalorespiration. Les molécules dégradées par co-métabolisme sont indiquées en gras.

Souches	Molécules organo-halogénées dégradées	Produits	Références
195	Chloroéthènes (PCE, TCE, 1,1-DCE, cis-DCE, trans-DCE , VC)	VC, Ethylène	Maymo-Gatell <i>et al.</i> , 1997; 2001
	Haloalcane (1,2-DCA, 1,2-DBA)	Ethane	
	Hexachlorobenzène (HCB)	Chlorobenzènes moins chlorés (1,2,3,5-TeCB + 1,3,5-TrCB)	
	PCDDs (1,2,3,4-TeCDD)	PCDDs moins chlorés (1,2,4-TrCDD + 1,3-DiCDD)	
	PCBs (2,3,4,5,6- Pentachlorobiphenyl)	PCBs moins chlorés (2,3,4,6- et/ou 2,3,5,6-tetrachlorobiphenyl + 2,4,6-trichlorobiphenyl)	Fennell <i>et al.</i> , 2004
	1,2,3,4-Tetrachloronaphtalène	dichloronaphtalène non identifié	
	Chlorophénols (2,3-dichlorophénol, 2,3,4-trichlorophénol, 2,3,6-trichlorophénol)	Chlorophénols moins chlorés (avec élimination du Cl en ortho)	Adrian <i>et al.</i> , 2007
	Ethers diphenyles Polybromés PBDEs (hexa, hepta, octa, nona-BDEs)	PBDEs moins bromés (penta, hexa, hepta-BDEs)	He <i>et al.</i> , 2006
BAV1	Chloroéthènes (PCE , TCE, 1,1-DCE, cis-DCE, trans-DCE, VC) et Bromoéthènes (VB)	Ethylène	He <i>et al.</i> , 2003
	Chloroalcane (1,2-DCA)	Ethane	
CBDB1	Chlorobenzènes (1,2,3-TCB, 1,2,4-TCB, 1,2,3,4-TeCB, 1,2,3,5-TeCB et 1,2,4,5-TeCB)	Chlorobenzènes moins chlorés (dichlorobenzènes ou 1,3,5-TCB)	Adrian <i>et al.</i> , 2000
	Chlorophénols (2,3-dichlorophénol, trichlorophénols, tetrachlorophénols, pentachlorophénol)	Chlorophénols moins chlorés	Adrian <i>et al.</i> , 2007
	PCDDs (1,2,3- or 1,2,4-TrCDD, 1,2,3-TrCDD, 1,2,4-TrCDD, 2,3-DiCDD, 1,2,3,4-TeCDD, 1,2,3,7,8-PeCDD)	PCDDs moins chlorés (1,3-DiCDD, 2,3-DiCDD, 2-MCDD, 2,3,7,8-TeCDD, 2,7-DiCDD, 2,8-DiCDD, 1,3,7,8-TeCDD, 2,3,7-	Bunge <i>et al.</i> , 2005
VS	Chloroéthènes (cis-DCE, VC)	Ethylène	Cupples <i>et al.</i> , 2003
FL2	Chloroéthènes (PCE , TCE, cis-DCE, trans-DCE, VC)	VC, Ethylène	He <i>et al.</i> , 2005
GT	Chloroéthènes (TCE, 1,1-DCE, cis-DCE, VC)	Ethylène	Sung <i>et al.</i> , 2006
DCMB5	PCDDs (1,2,4-TrCDD)	PCDDs moins chlorés (1,3-DiCDD + 2-MCDD)	Bunge <i>et al.</i> , 2008
	Chlorobenzène (1,2,3-TCB)	Chlorobenzène moins chlorés (1,3-DCB)	
MB	Chloroéthènes (PCE, TCE)	trans-DCE	Cheng and He, 2009
	Ethers diphenyles polybromés PBDEs (octa-BDEs)	BDEs moins bromés (penta, tétra-BDEs)	
ANAS1	Chloroéthènes (TCE, 1,1-DCE, cis-DCE)	VC	Lee <i>et al.</i> , 2011
ANAS2	Chloroéthènes (TCE, 1,1-DCE, cis-DCE, VC)	Ethylène	

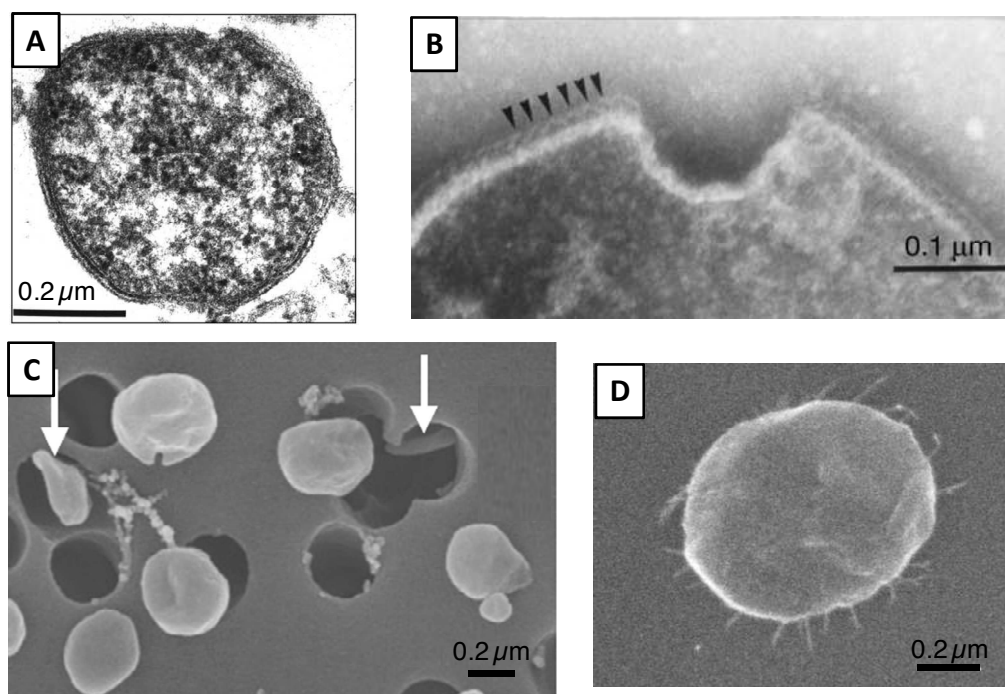


Figure 13 : Clichés de microscopie électronique de plusieurs souches de *Dehalococcoides*. A : *Dehalococcoides ethenogenes* 195 (Maymo-Gatell *et al.*, 1997). B : *Dehalococcoides* sp. CBDB1, les flèches indiquent la paroi particulière dépourvue de peptidoglycane (Adrian *et al.*, 2000). C : *Dehalococcoides* sp. FL2, les individus indiqués par les flèches illustrent la forme discoïdale de ces cellules (He *et al.*, 2005). D : *Dehalococcoides* sp. BAV1 (He *et al.*, 2003).

al., 2000), VS (Cupples *et al.*, 2003; Muller *et al.*, 2004), BAV1 (He *et al.*, 2003), FL2 (He *et al.*, 2005), GT (Sung *et al.*, 2006a), DCMB5 (Bunge *et al.*, 2008), MB (Cheng and He, 2009), ANAS1 et ANAS2 (Lee *et al.*, 2011). Les souches de *Dehalococcoides* sont d'excellents modèles pour comprendre l'évolution de la diversification de la déhalorespiration. En effet, bien qu'elles soient phylogénétiquement très proches, comme en atteste l'identité de leur séquences ARNr 16S >98%, elles sont capables d'utiliser une très large variété de composés organohalogénés (Tableau 8).

3.1.1. Caractéristiques morphologiques et physiologiques

Ce groupe *Dehalococcoides* est unique au regard de certaines caractéristiques comme la forme et la taille de ses cellules (petits discoïdes plutôt que des coccoïdes, comme décrit initialement par Maymo-Gatell *et al.* en 1997, d'environ 0,3 à 1 μm de diamètre). La paroi cellulaire de ces dernières est dépourvue de peptidoglycane et présente une ultrastructure similaire à celle de la couche de surface (couche-S) de certaines *Archaea* (Figure 13), ce qui leur confère ainsi une résistance aux antibiotiques agissant sur la synthèse du peptidoglycane (β -lactamines et vancomycine). Leur mode de respiration strict les distingue également du fait de l'utilisation unique de l' H_2 comme donneur d'électrons (hydrogénotrophes stricts) et des composés organohalogénés comme accepteurs finaux d'électrons. Les *Dehalococcoides* sont des bactéries mésophiles (25-40°C) qui préfèrent des pH neutres. Leur maintien en culture pure est très difficile du fait de leur croissance très lente (temps de doublement compris entre 19,2 heures pour *D. ethenogenes* 195 et 2 à 2,5 jours pour la souche GT), de leur forte sensibilité à l'oxygène (Amos *et al.*, 2008) et à la température (Fletcher *et al.*, 2011), mais aussi de leurs besoins en co-facteurs souvent apportés par d'autres microorganismes (cobalamine par exemple) (Tas *et al.*, 2010).

3.1.2. Caractéristiques génomiques

En plus de leurs caractéristiques morphologiques et physiologiques particulières, les *Dehalococcoides* présentent des caractéristiques étonnantes au niveau génomique. En effet, leur génome de forme circulaire est de taille très réduite, de l'ordre de 1,5 Mpb (Tableau 9) et compte donc parmi les plus petits génomes pour des bactéries libres (Kube *et al.*, 2005; McMurdie *et al.*, 2009; Tas *et al.*, 2010).

Tableau 9 : Taille et composition des génomes des souches déhalorespirantes obligatoires appartenant aux genres *Dehalococcoides* et *Dehalogenimonas*.

Genre	Souche	Taille Génome (Kpb)	ORFs	%GC	Nombre de gènes <i>rdhA</i>	Références
<i>Dehalococcoides</i>	<i>ethenogenes</i> 195	1469	1647	48,9	18	Seshadri <i>et al.</i> , 2005
	CBDB1	1395	1517	47	32	Kube <i>et al.</i> , 2005
	BAV1	1341	1443	47	10	McMurdie <i>et al.</i> , 2009
	VS	1413	1489	47,3	36	
	GT	1360	1483	47	20	
<i>Dehalogenimonas</i>	BL-DC-9 ^T	1686	1771	55	19	non publié

Kpb : kilo paires de bases, ORF : Open Reading Frames, *rdhA* : gène codant pour une déhalogénase réductrice

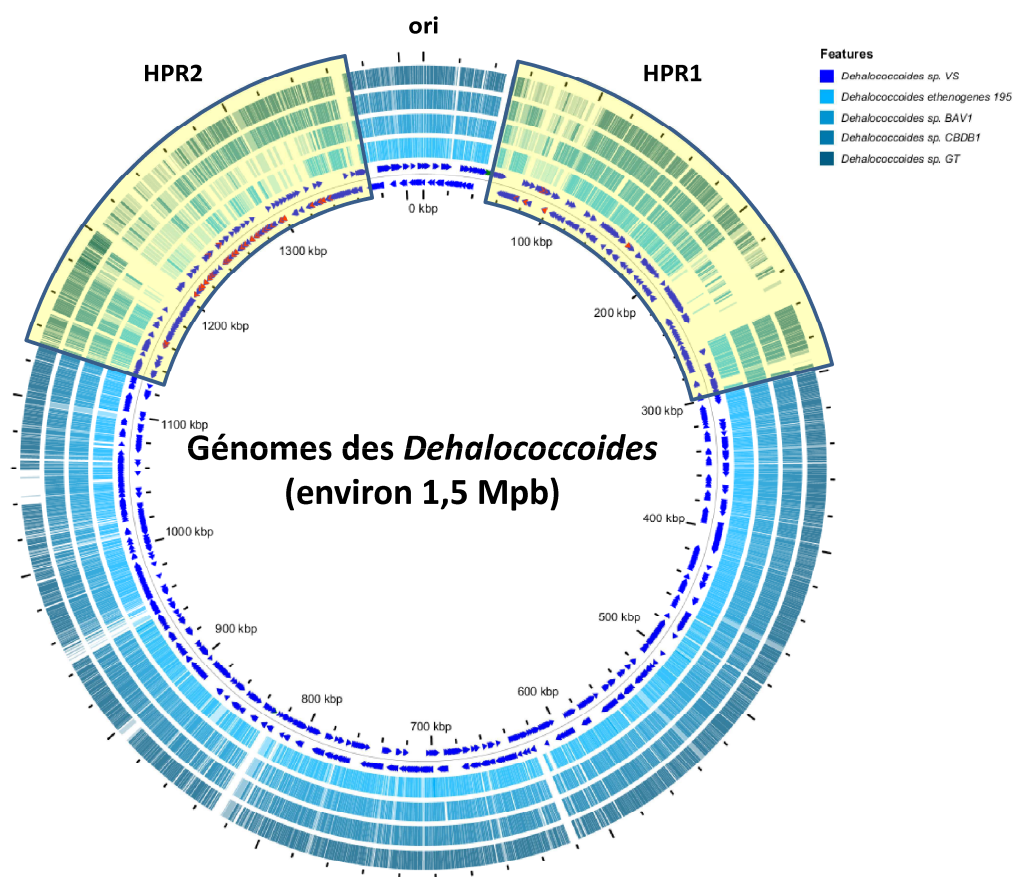


Figure 14 : Comparaison des génomes de cinq souches de *Dehalococcoides* (souches VS, *ethenogenes* 195, BAV1, CBDB1 et GT) réalisée à l'aide du logiciel GView (Petkau *et al.*, 2010) en utilisant comme référence le génome de la souche VS. Les cercles représentent (du centre vers l'extérieur) : 1-2, le génome de la souche VS avec les deux brins montrant la position des CDSs en bleu, les gènes codant pour des RdhA putatifs sont visualisés en rouge ; 3-6, les génomes des souches *ethenogenes* 195, BAV1, CBDB1 et GT respectivement. L'intensité de coloration pour chaque souche reflète le degré de conservation entre les orthologues des différentes souches (forte intensité = forte conservation). Les deux régions de forte plasticité (HPR1 et 2) sont surlignées en jaune. ori : origine de réplication.

Actuellement, les séquences complètes de cinq génomes de *Dehalococcoides* sont disponibles : *D. ethenogenes* 195 (Seshadri *et al.*, 2005), CBDB1 (Kube *et al.*, 2005), GT (non publié), VS et BAV1 (McMurdie *et al.*, 2009). Quel que soit le génome auquel on s'intéresse, ces derniers ne disposent que d'une seule copie des gènes codant pour les ARN ribosomiaux (16S, 23S et 5S). De plus, ils contiennent un nombre de gènes assez restreint, estimé en moyenne à 1500. Aucun plasmide n'a été décrit à ce jour. Le mode de vie extrêmement spécialisé de ces microorganismes, puisqu'ils occupent une niche physiologique unique (déhalorespiration de molécules organohalogénées avec comme seul donneur d'électrons l'H₂), est certainement à l'origine de cette réduction.

En revanche, l'annotation de ces génomes a révélé une grande richesse en gènes codant des déhalogénases réductrices putatives (*rdhA*), avec un maximum de 36 gènes pour la souche VS et un minimum de 10 pour BAV1 (Tableau 9). Ces résultats confirment ce qui avait déjà été supposé au travers des travaux de Hölscher *et al.* (2004) par des approches plus classiques d'amplification PCR. Une telle diversité en gènes *rdhA* permet d'expliquer les capacités d'adaptation étonnantes des souches affiliées au genre *Dehalococcoides*, leur offrant ainsi une grande flexibilité de leur système respiratoire et la possibilité d'utiliser une large variété de molécules halogénées comme accepteurs finaux d'électrons (McMurdie *et al.*, 2009). La comparaison de ces différents génomes (Kube *et al.*, 2005; McMurdie *et al.*, 2009) révèle également une structure constituée :

- d'un « core genome » très stable représentant 68 à 77 % du génome et contenant tous les gènes du métabolisme central. Ces gènes ont conservé un fort degré de similarité au niveau des séquences nucléiques mais aussi de la synténie (ordre des gènes, orientation et contexte génomique) (Figure 14) ;

- de deux régions de forte plasticité (HPRs : High Plasticity Regions), de taille pouvant atteindre 200 Kpb, situées de part et d'autre de l'origine de réplication (ori). Elles sont caractérisées par la présence de séquences impliquées dans les mécanismes de recombinaisons (séquences d'insertions, de délétions et d'inversions, éléments transposables, recombinaisons, intégrases, séquences phagiques) et pourraient donc être le siège de réarrangements génomiques importants. La grande majorité des CDSs (Coding DNA Sequences) codant pour des déhalogénases réductrices sont, d'une part, localisés dans ces régions (91 des 96 gènes retrouvés dans les génomes des souches 195, CBDB1, VS et BAV1) suggérant ainsi une forte mobilité de ces gènes, et d'autre part, présentent un biais d'orientation important puisqu'ils sont quasiment tous portés par le brin direct (Kube *et al.*, 2005; Seshadri *et al.*, 2005; McMurdie *et al.*, 2009). Ce biais étant généralement observé pour les gènes « essentiels »

chez les bactéries (Rocha and Danchin, 2003a, b; Lin *et al.*, 2010), cette orientation particulière souligne l'importance des déhalogénases réductrices chez les *Dehalococcoides*.

3.1.3. Transferts horizontaux de gènes chez *Dehalococcoides*

Grâce à la disponibilité de la séquence de plusieurs génomes de souches affiliées au genre *Dehalococcoides*, il a été possible de réaliser des analyses approfondies des séquences des gènes codant pour les RdhA ainsi que de leur environnement génétique. C'est de cette façon que Regear et ses collaborateurs ont pu apporter des éléments indiquant qu'au moins 15 des 18 gènes codant pour des RdhA putatives présents chez la souche *D. ethenogenes* 195 ont été acquis par transfert horizontal (Regear *et al.*, 2005). Ces résultats ont par la suite été corroborés par l'étude des régions chromosomiques contenant l'opéron *tceAB* de plusieurs souches environnementales de *Dehalococcoides* provenant de sites contaminés distants d'un point de vue géographique (Krajmalnik-Brown *et al.*, 2007). En effet, ils ont pu identifier chez certaines souches la présence, en aval du gène *tceB*, d'une séquence putative d'élément transposable, séquence qui serait à l'origine de l'acquisition de la totalité de l'opéron contenant ces gènes.

L'analyse des séquences des gènes codant pour les VC-réductases (*vcrA* et *bvcA*), responsables de la déchloration du DCE et du VC, a également conduit à une découverte étonnante. En effet, même si seuls des membres appartenant au groupe des *Dehalococcoides* sont actuellement capables de respirer ces composés, leur gènes pourraient avoir été acquis par transfert horizontal à partir de souches n'appartenant probablement pas à ce genre (McMurdie *et al.*, 2009). Cette hypothèse repose sur (i) l'observation d'un fort biais d'usage des codons pour les gènes *vcrA* et *bvcA* (McMurdie *et al.*, 2007) qui n'est observé pour aucun autre gène présent sur les génomes de *Dehalococcoides*, et (ii) la localisation de ces gènes dans des îlots génomiques de faible pourcentage en GC qui interrompent la conservation de synténie locale des gènes également présents chez les autres souches de *Dehalococcoides* (McMurdie *et al.*, 2009).

Ainsi, toutes ces observations suggèrent fortement que les bactéries du genre *Dehalococcoides* sont en perpétuelle évolution et qu'elles acquièrent de nouvelles capacités de dégradation par transferts horizontaux de gènes, non seulement entre membres de ce groupe mais également avec des bactéries appartenant à d'autres genres. Ceci leur permet ainsi de s'adapter rapidement à de nouvelles molécules halogénées et de nouvelles conditions de milieux.

3.2. Les autres groupes déhalorespirants affiliés aux *Chloroflexi*

3.2.1. Le genre *Dehalogenimonas*

Le genre *Dehalogenimonas*, qui appartient à la classe des *Dehalococcoidetes*, ne se compose, à ce jour, que d'une seule espèce, *Dehalogenimonas lykanthroporepellens*, avec deux souches, BL-DC-8^{PP} et BL-DC-9^T (Moe *et al.*, 2009). Isolées d'une nappe phréatique américaine contaminée, ces dernières sont capables d'utiliser une grande variété d'alcane polychlorés par le processus de déhalorespiration, avec comme donneur d'électrons l'H₂ (Yan *et al.*, 2009a). En revanche, elles sont incapables de déhalogéner les chloroéthènes. Leur affiliation à un nouveau genre est basée sur leurs séquences ARNr 16S qui ne présentent environ que 90% d'identité avec celles des souches du genre *Dehalococcoides*. Seul le génome de la souche BL-DC-9^T est actuellement disponible (non publié). Sa taille, proche de 1,7 Mpb, et son pourcentage GC de 55%, sont légèrement plus élevés que ce qui est observé pour les souches de *Dehalococcoides*. En revanche, il contient également un grand nombre de gènes codant pour des RdhA putatives, avec 19 séquences identifiées (Tableau 9).

3.2.2. Le groupe des *Dehalococcoides*-like (DLG)

Le troisième groupe de microorganismes déhalorespirants, appartenant aux *Chloroflexi*, mais non classé parmi les *Dehalococcoidetes*, est couramment appelé « *Dehalococcoides*-like group » (DLG). Il ne contient, pour l'instant, que des souches environnementales encore non affiliées à un niveau taxonomique plus fin. Les membres de ce groupe présentent une large distribution dans l'environnement comme cela a pu être montré par des approches d'enrichissement en microcosmes, ces derniers étant réalisés à partir d'échantillons provenant de milieux variés (sédiments marins, rivières, lacs et estuaires) et contaminés par diverses molécules organochlorées (PCBs ou PCDD/Fs) (Fagervold *et al.*, 2005; Yoshida *et al.*, 2005; Yan *et al.*, 2006; Ahn *et al.*, 2007; Fagervold *et al.*, 2007; Ahn *et al.*, 2008). Certains ont, par ailleurs, des capacités de dégradation très intéressantes. Par exemple, la souche o-17, isolée des sédiments de l'estuaire de Chesapeake Bay aux Etats-Unis et maintenue en culture mixte (Cutter *et al.*, 2001), dégrade plusieurs PCBs mais également des chlorobenzènes et certains chloroéthènes dont le PCE et le TCE (May *et al.*, 2006). Une seconde souche, avec des capacités similaires mais provenant de sédiments d'un autre estuaire américain (Wu *et al.*, 2000; Wu *et al.*, 2002b; Wu *et al.*, 2002a; Miller *et al.*, 2005), a quant à

elle été isolée en culture pure. Elle a alors été dénommée *Dehalobium chlorocoercia* DF-1 (May *et al.*, 2008). Même si ces deux souches sont physiologiquement proches des *Dehalococcoides*, leurs séquences ARNr 16S sont assez éloignées (89 et 87 % d'identité respectivement avec les *Dehalococcoides*).

Durant la dernière décennie, les approches moléculaires ont permis d'identifier de nombreuses séquences ARNr 16S proches de celles de microorganismes déhalorespirants affiliés aux *Chloroflexi* ou de gènes codant des RdhA encore non-identifiées à partir d'environnements contaminés par des composés organohalogénés d'origine diverse (Hiraishi *et al.*, 2005; Watts *et al.*, 2005; Ahn *et al.*, 2007; Kittelmann and Friedrich, 2008; Futagami *et al.*, 2009). Ces résultats suggèrent l'existence de microorganismes qui pourraient être porteurs de nouvelles fonctions permettant de dégrader une plus large gamme de composés chimiques parmi lesquels, potentiellement, des chlorés qui sont considérés comme récalcitrants. Ils constituent donc une ressource exploitable importante pour de nouvelles applications en bioremédiation afin de traiter plus efficacement certains sites contaminés.

Tableau 10 : Répartition approximative des volumes d'eau douce dans les différents compartiments terrestres. La glace n'est pas prise en compte (adapté du document « Vulnérabilité des aquifères et évaluation de la pollution » de l'Agence Internationale de l'Energie Atomique, <http://www-naweb.iaea.org/>).

Type d'eau	Pourcentage	Volume
Eau souterraine	96,3%	8 000 000 km ³
Eau de lac	2,7%	226 000 km ³
Humidité du sol	0,8%	62 000 km ³
Humidité de l'air	0,2%	15 000 km ³
Eau de rivière	0,01%	1 000 km ³

4. Stratégies de réhabilitation des aquifères contaminées par les solvants chlorés

L'eau, longtemps considérée comme une ressource inépuisable, est devenue aujourd'hui une ressource fragile du fait, notamment, des besoins grandissants d'une population mondiale en plein essor, du manque chronique enduré par certains pays, et des pollutions croissantes des réserves, en particulier, dans les pays industrialisés. Les réserves mondiales en eau des nappes souterraines représentent presque 97 % de toute l'eau douce disponible sur les continents (document « Vulnérabilité des aquifères et évaluation de la pollution » de l'Agence Internationale de l'Energie Atomique, <http://www-naweb.iaea.org/>) (Tableau 10). En outre, les nappes alimentent en eau douce le réseau superficiel des cours d'eau. Elles doivent donc impérativement être protégées contre les sources de pollution diffuse, d'autant que le renouvellement de ces eaux souterraines peut être très lent et que les pollutions y sont persistantes.

Dans ce contexte et afin de disposer de façon durable d'une eau de bonne qualité, la préservation des nappes phréatiques fait partie des enjeux économiques et sociétaux incontournables du XXIème siècle. Parmi les mesures mises en place figure l'instauration de périmètres de protection autour des captages. Celle-ci est rendue obligatoire par la législation de nombreux pays mais elle n'offre qu'une protection limitée. En effet, une pollution éloignée peut gagner, même diluée, ces eaux protégées plusieurs années après l'événement polluant. Il faut donc également diminuer les sources de pollution ce qui passe, notamment, par la restauration des sites contaminés. Pour cela, les approches de remédiation ont été développées. Elles consistent à éliminer les sources de pollution et à maîtriser les panaches de composés dissous.

Les premières prises de conscience des problèmes liés à la qualité de l'eau datent des années 1980s et ont donné lieu, en particulier aux Etats-Unis, à l'adoption de lois engageant la responsabilité conjointe des industriels et des autorités publiques pour encadrer la protection de l'environnement, la restauration des sites contaminés et la conservation des ressources (CERCLA pour « Comprehensive Environmental Response, Compensation, and Liability Act », acté par le congrès américain le 11 Décembre 1980). Ceci s'est immédiatement traduit par le démarrage du programme « Superfund » par l'Environmental Protection Agency (EPA). A l'époque, les actions de remédiation consistaient principalement au pompage de l'eau contaminée suivie du traitement et/ou du stockage de cette dernière dans des lieux

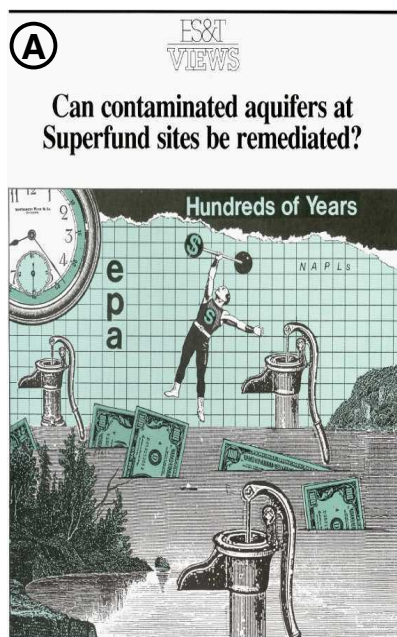


Figure 15 : Premières pages d'articles parus dans les années 1990 dans la revue Environmental Science and Technology. A : Travis and Doty, 1990. B : McDonald and Kavanaugh, 1994.

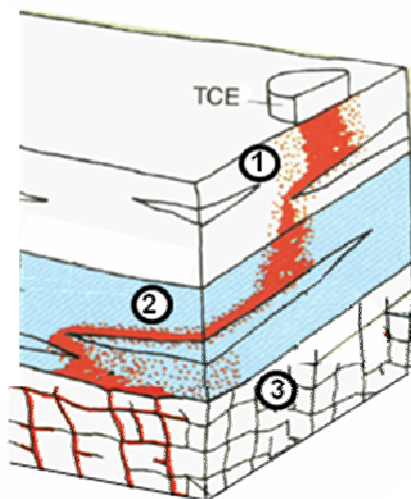


Figure 16 : Représentation schématique des principales contraintes rencontrées pour la mise en place de stratégies de remédiation efficaces après pollution d'une nappe phréatique par les chloroéthènes. 1 : formation de gouttelettes piégées entre les particules de sol de la zone non-saturée. 2 : hétérogénéité des aquifères. 3 : diffusion des contaminants dans des zones inaccessibles comme les fractures de roche (d'après Mackay et Cherry, 1989).

sécurisés (actions regroupées sous le terme générique « Pump and Treat ») (Mackay and Cherry, 1989). Cependant, dès les premières applications, les acteurs ont réalisé les difficultés de mise en œuvre de cette stratégie provoquant alors, au sein de la communauté scientifique, le doute sur l'efficacité et la faisabilité de tels traitements (Figure 15) (Travis and Doty, 1990; MacDonald and Kavanaugh, 1994).

Les principales difficultés rencontrées concernaient : (i) l'hétérogénéité des aquifères constitués de nombreuses couches de composition différente (sable, gravier, roche, argile...), ce qui compliquait fortement l'élaboration de stratégies efficaces ; (ii) la migration de DNAPLs jusqu'au fond de l'aquifère entraînant la formation de gouttelettes piégées entre les particules de sol de la zone non-saturée, les rendant donc impossibles à repérer et à éliminer ; (iii) la diffusion des contaminants dans des zones inaccessibles, par exemple dans les fractures de roche ou les pores de certaines couches géologiques, source d'un relargage de composés dissous dans la nappe ; (iv) les difficultés à bien caractériser les zones sources de contamination, ce qui nécessite une connaissance approfondie de l'historique de pollution et des propriétés hydrogéologiques du site (Figure 16) (Mackay and Cherry, 1989; MacDonald and Kavanaugh, 1994).

La dépollution des nappes phréatiques par pompage peut donc s'avérer délicate pour de nombreux sites, nécessitant alors des dizaines voire des centaines d'années sans garantie de succès. Par conséquent, considérée comme un procédé de traitement « à perpétuité » (Mackay and Cherry, 1989), cette solution est apparue comme inadaptée pour permettre la restauration des aquifères, que ce soit sur le plan de l'efficacité de la méthode (retrouver des teneurs limites à ne pas dépasser en polluants dissous pour que l'eau soit conforme aux normes fixées par les autorités sanitaires et cela dans un délai raisonnable) ou sur le plan économique (Travis and Doty, 1990). Il a donc fallu développer de nouvelles approches basées sur des processus physiques, chimiques ou biologiques afin de répondre à ces contraintes et permettre une restauration réellement efficace des aquifères quelle que soit la configuration du site (Henry *et al.*, 2002; Khan *et al.*, 2004).

4.1. Approches physiques

Ces techniques exploitent les propriétés physiques des molécules polluantes afin de réaliser leur extraction de la zone contaminée. Concernant les chloroéthènes, plusieurs stratégies ont été mises au point. Elles s'appuient le plus souvent sur la volatilité, la solubilité et la mobilité de ces composés. Cependant, elles peuvent également reposer sur l'injection de

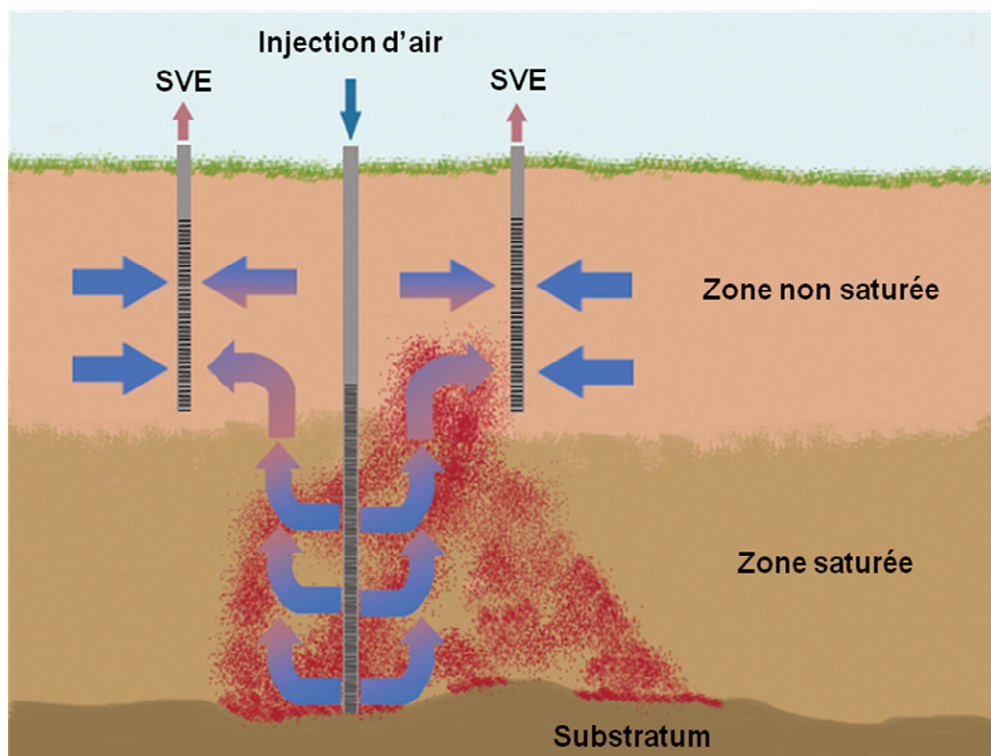


Figure 17 : Principe de la technique d'air sparging (d'après Christ *et al.*, 2005). SVE : système d'extraction de vapeurs.

produits tensioactifs (surfactants) ou de solvants afin de modifier leurs propriétés et ainsi faciliter leur extraction (Henry *et al.*, 2002). Parmi ces approches figurent :

- **L'air sparging** : cette technique est largement utilisée pour traiter les aquifères contaminés par des composés organiques volatils (COVs) dont font partie les chloroéthènes (Bass *et al.*, 2000; Benner *et al.*, 2002; Adams *et al.*, 2011). Elle consiste à injecter de grosses quantités d'air sous pression dans la nappe phréatique *via* un ou plusieurs puits d'injection ce qui provoque son aération et la volatilisation des polluants vers la zone non-saturée (vadose). Elle est très souvent couplée à un système d'extraction de vapeurs (SVE). Pour cela, plusieurs puits verticaux et/ou horizontaux sont forés pour collecter les polluants présents dans le sol à l'état gazeux (Figure 17) (Khan *et al.*, 2004; Christ *et al.*, 2005). L'air sparging est une technique assez efficace pour traiter les chloroéthènes puisqu'elle permet d'éliminer la quasi-totalité des polluants (>90%) sur la plupart des sites (Bass *et al.*, 2000). Cependant, dans quelques cas, cette efficacité n'est que partielle et un retour de fortes concentrations en polluants dissous dans la nappe peut être observé après l'arrêt du traitement. Les durées de traitement vont de quelques mois jusqu'à deux ans, et les coûts sont variables. Ils dépendent surtout de l'importance de la contamination et de l'étendue de la zone à traiter. L'utilisation de cette technique doit être pratiquée avec précaution, notamment lorsque d'importantes DNAPLs contaminent la nappe phréatique, afin d'éviter que les injections ne provoquent l'extension de la zone polluée par dispersion du polluant (Henry *et al.*, 2002; Christ *et al.*, 2005). De manière indirecte, l'air sparging entraîne l'oxygénation de la nappe phréatique ce qui peut stimuler les processus aérobies de biodégradation des chloroéthènes (Bass *et al.*, 2000).

- **Le traitement thermique** : par ce procédé, la température au sein de la zone contaminée est augmentée ce qui favorise la désorption et la volatilisation des polluants, la pression de vapeur d'un composé doublant à peu près tous les 10°C (Henry *et al.*, 2002). Cette approche est très efficace pour éliminer les COVs. Pour les chloroéthènes, l'augmentation de la température est obtenue grâce à l'injection de vapeur d'eau (« Steam Stripping ») (Kaslusky and Udell, 2002; Gudbjerg *et al.*, 2004) ou par d'autres méthodes plus originales comme l'utilisation de microondes (Acierno *et al.*, 2003; Robinson *et al.*, 2008) ou d'électrodes (Electrical Resistance Heating), ces dernières induisant alors un fort courant électrique dans le milieu pollué (Friis *et al.*, 2007; Truex *et al.*, 2011). Ces traitements, provoquant la volatilisation des polluants, doivent être couplés à un système SVE afin de récupérer les vapeurs toxiques. La demande énergétique importante nécessaire à la mise en œuvre de ces traitements engendre des coûts conséquents qui limitent leur utilisation à grande

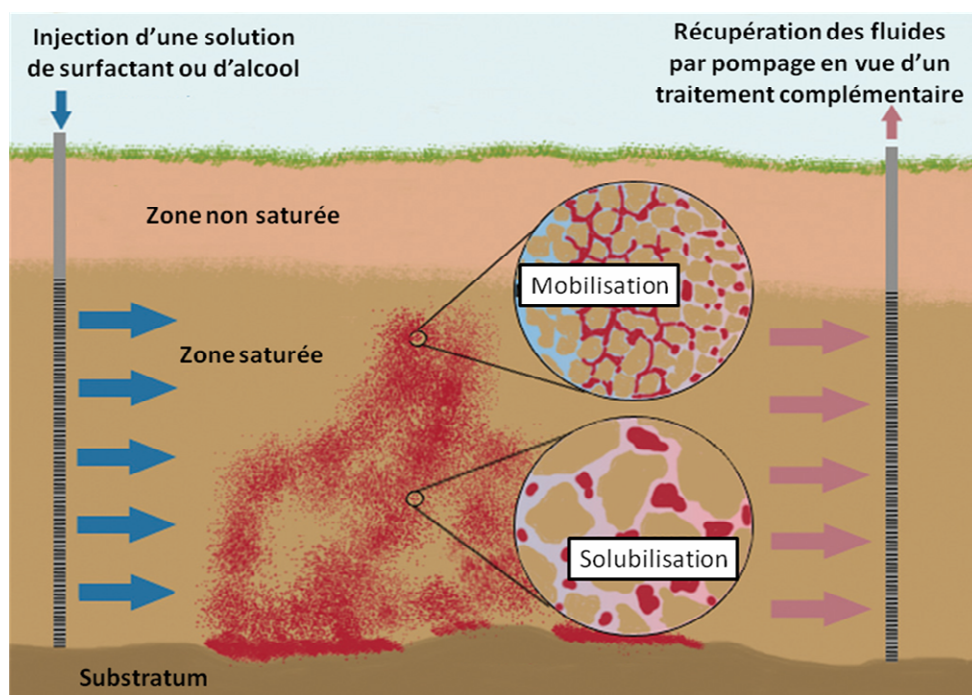


Figure 18 : Utilisation de surfactants ou d'alcools pour traiter les aquifères contaminés par des chloroéthènes (d'après Christ *et al.*, 2005).

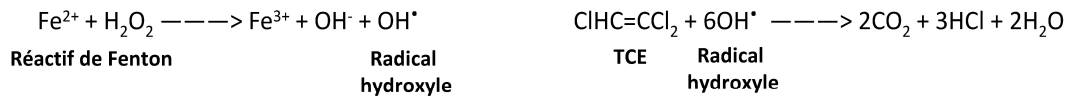
échelle (Christ *et al.*, 2005). De plus, comme pour la technique d'air sparging, le traitement thermique peut engendrer, si il n'est pas correctement contrôlé, une mobilisation et une dispersion sous l'effet de la chaleur des polluants et de leurs sous-produits de dégradation (extension de la zone contaminée). Enfin, malgré le peu d'études disponibles à l'heure actuelle, l'impact d'un tel traitement sur la macrofaune et la microflore hébergées par l'écosystème pourrait être assez néfaste (Krauter *et al.*, 1995; Richardson *et al.*, 2002; Christ *et al.*, 2005).

- **L'injection de tensioactifs (surfactants) ou de co-solvants** : cette technique est utilisée en complément de l'approche classique par pompage car elle favorise la mobilité, la solubilité et la désorption des polluants (Figure 18) (Henry *et al.*, 2002; Christ *et al.*, 2005). Les agents tensioactifs sont des molécules amphiphiles. Elles présentent donc une partie lipophile et une autre hydrophile, ce qui est nécessaire pour solubiliser deux phases non miscibles. Il en existe une grande variété mais les surfactants anioniques (exemple : SDS), et non-ioniques (exemples : TritonX100, Tween80) sont les mieux adaptés pour traiter les sites contaminés par les chloroéthènes (Knox *et al.*, 1997; Zhao *et al.*, 2006). Puisque ces polluants sont également solubles dans les alcools (en particulier l'éthanol), ces derniers peuvent être utilisés comme co-solvants pour les solubiliser et faciliter leur extraction par pompage (Jawitz *et al.*, 2000). L'utilisation de cette approche nécessite d'être contrôlée pour éviter toute propagation de la contamination hors de la zone traitée (Christ *et al.*, 2005). Par ailleurs, l'effet de l'injection de ces différents composés sur la microflore reste encore aujourd'hui assez peu documenté. Les résultats des premières études montrent qu'elles auraient plutôt un effet bénéfique. En effet, leur dégradation pourrait permettre de produire des composés plus simples comme des acides gras volatils ou de l'H₂ directement assimilables par les populations capables de dégrader les chloroéthènes (Mravik *et al.*, 2003; Ramsburg *et al.*, 2004).

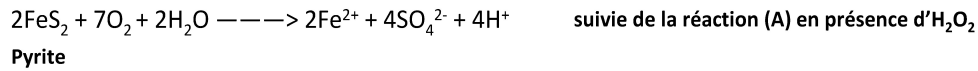
4.2. Approches chimiques

Sous cette appellation sont regroupées toutes les stratégies faisant intervenir des réactions chimiques permettant de détruire ou de modifier des molécules polluantes. Quand elles sont appliquées directement *in situ*, leur effet est en général très rapide (Henry *et al.*, 2002). Les deux types de réactions chimiques pouvant être utilisées pour le traitement des chloroéthènes sont l'oxydation ou la réduction. L'oxydation de ces derniers peut être obtenue au cours de plusieurs réactions faisant intervenir différents composés comme le réactif de

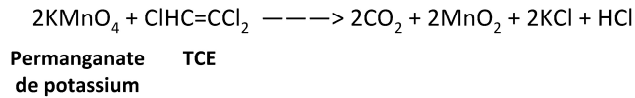
(A) Oxydation du TCE en présence de réactif de Fenton



(B) Oxydation du TCE en présence de pyrite



(C) Oxydation du TCE en présence de permanganate de potassium (ou de sodium)



(D) Oxydation du TCE en présence d'anion persulfate et d'un activateur (ici le Fe^{2+})

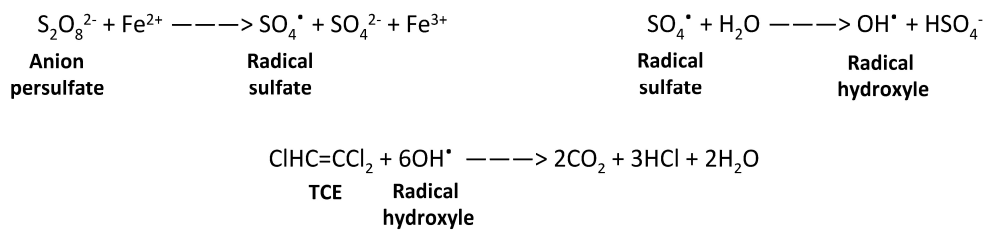


Figure 19 : Réactions chimiques permettant l'oxydation du TCE.

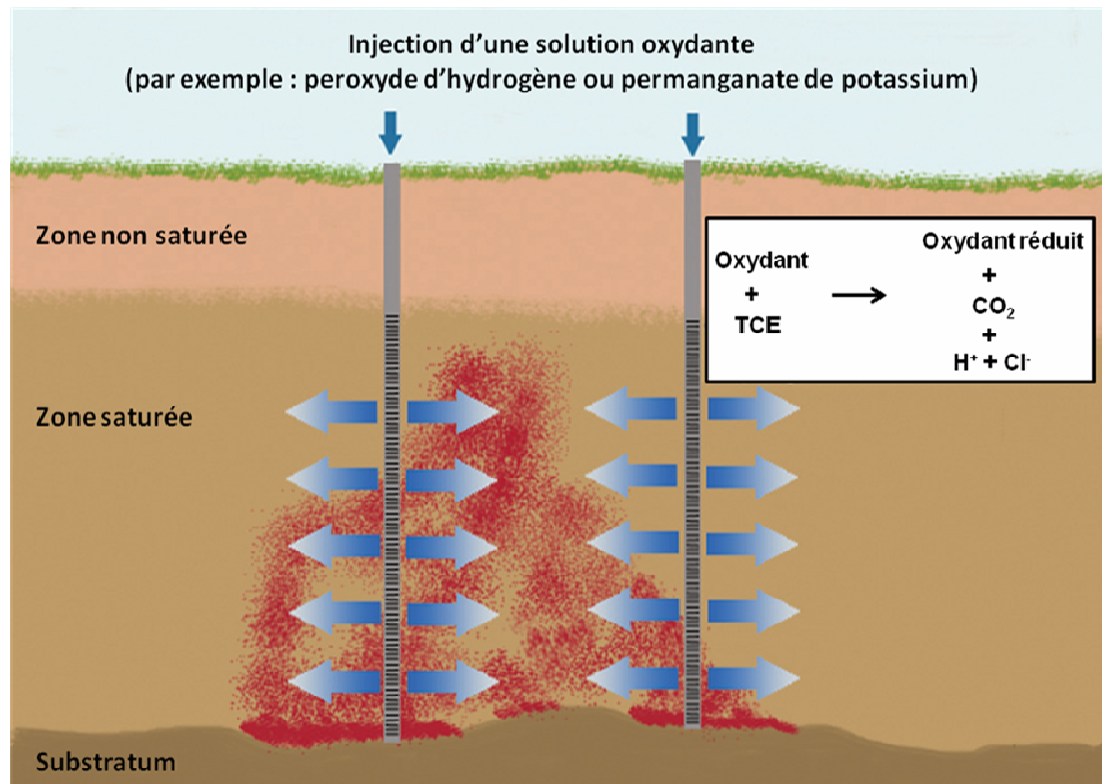


Figure 20 : Technique de remédiation chimique par injection directe d'une solution oxydante dans la nappe phréatique (d'après Christ *et al.*, 2005).

Fenton (radical hydroxyle produit lors de la réaction entre le peroxyde d'hydrogène H_2O_2 et le Fe^{II}) (Chen *et al.*, 2001), la pyrite (FeS_2) (Che *et al.*, 2011), le permanganate de sodium (NaMnO_4^-) ou de potassium (KMnO_4^-) (Schroth *et al.*, 2001; Kao *et al.*, 2008), ou encore l'anion persulfate ($\text{S}_2\text{O}_8^{2-}$) (Liang *et al.*, 2008) (Figure 19). Elles conduisent à la minéralisation totale des chloroéthènes en CO_2 et en sels (NaCl , KCl ...). La réduction (ou déchloration réductrice), quant à elle, est réalisée grâce à des métaux fortement réduits comme le Fe^0 (Roberts *et al.*, 1996; Quinn *et al.*, 2005) ce qui produit de l'éthylène ou de l'éthane.

Le traitement *in situ* par des approches chimiques peut être mis en place soit par injection directe de la solution contenant le réactif dans l'aquifère (Figure 20) (Quinn *et al.*, 2005) soit par l'élaboration de barrières réactives passives qui, placées en aval hydraulique de la zone source de pollution, utilisent le mouvement naturel de la nappe phréatique pour permettre le contact entre le polluant et le réactif (Figure 21) (Gavaskar, 1999). Une faible efficacité des techniques chimiques peut cependant être observée sur certains sites à cause de réactions compétitrices. En effet, il n'est pas exclu qu'en fonction de la nature de l'aquifère, le réactif agisse non seulement sur le polluant mais aussi sur d'autres composés naturels (Henry *et al.*, 2002).

4.3. Approches biologiques

Les stratégies physico-chimiques appliquées *in situ* pour la réhabilitation des aquifères contaminés par les chloroéthènes nécessitent la construction de dispositifs de traitements complexes (nombreux forages servant de puits de prélèvement, d'injection, ou de suivi ; implantation de systèmes d'extraction de vapeurs ou de barrières etc...), ce qui les rend donc particulièrement lourdes, onéreuses et invasives pour les écosystèmes traités. C'est pourquoi, les approches basées sur l'utilisation de systèmes biologiques pour éliminer les polluants industriels que ce soit en les dégradant, en les séquestrant ou en facilitant leur élimination (par exemple par volatilisation), sont des alternatives avantageuses aussi bien sur le plan économique qu'écologique (Perelo, 2010). Ces approches de remédiation biologique, également appelée bioremédiation, utilisent généralement des végétaux (phytoremédiation) ou des microorganismes (bioremédiation microbienne), voire même les deux en association, pour réduire la pression exercée par les substances polluantes sur l'environnement.

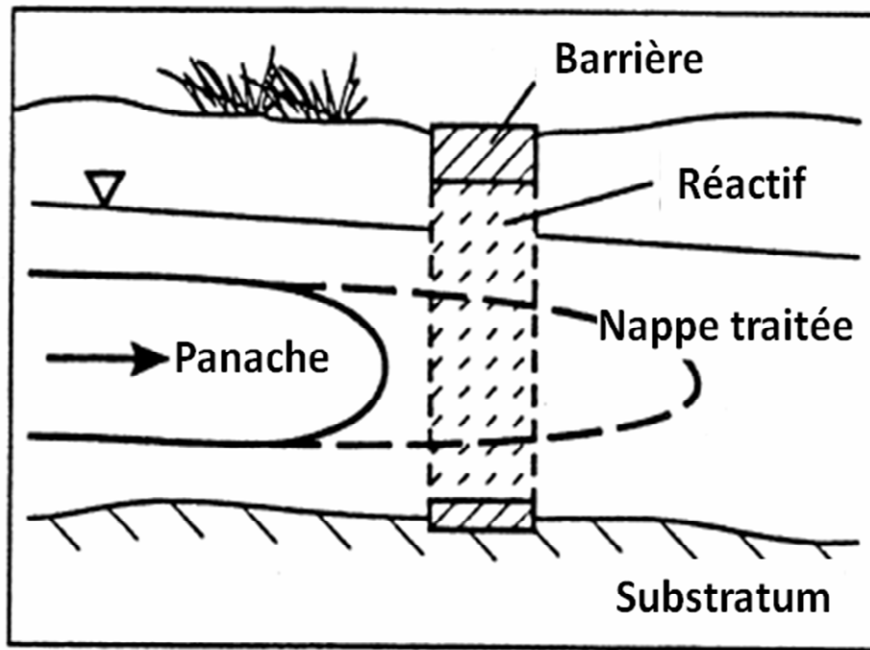


Figure 21 : Principe des barrières réactives passives placées en aval de la zone source de contamination (d'après Gavaskar, 1999).

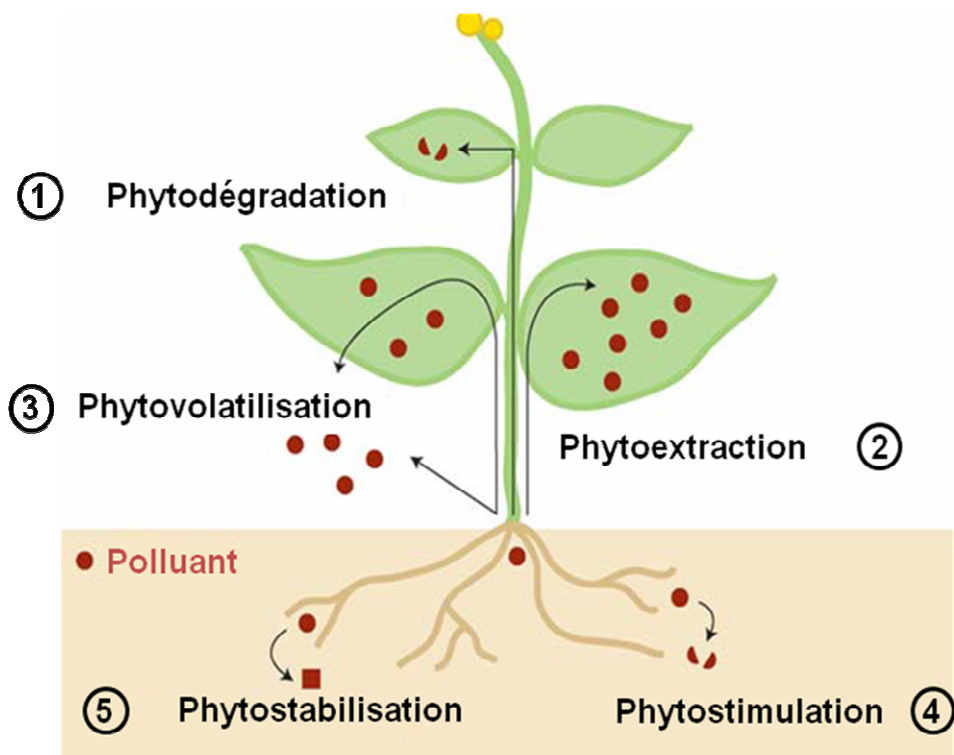


Figure 22 : Représentation schématique des différentes méthodes de phytoremédiation. Les polluants, représentés par des cercles rouges, peuvent être : 1) dégradés par la plante (phytodégradation), 2) séquestrés dans les parties aériennes de la plante (phytoextraction), 3) transpirés par la plante et rejetés dans l'atmosphère (phytovolatilisation), 4) dégradés par les microorganismes de la rhizosphère (phytostimulation), ou 5) stabilisés au niveau du système racinaire de la plante (phytostabilisation).

4.3.1. La phytoremédiation

Sous le vocable « phytoremédiation » (du grec *phyton* signifiant « végétal » et du latin *remedium* signifiant « rétablissement de l'équilibre ») sont regroupées toutes les technologies faisant intervenir les plantes et leur microflore associée pour éliminer, extraire, contenir ou rendre moins toxiques les contaminants présents dans l'environnement (Pilon-Smits, 2005). Il existe différentes méthodes pour réduire les niveaux de pollution (Figure 22) (Henry *et al.*, 2002; Pilon-Smits, 2005), à savoir :

- **La phytostimulation ou rhizodégradation :** l'élimination des contaminants se fait à travers l'activité microbienne présente dans la rhizosphère. Elle est stimulée par les exsudats racinaires libérés par les plantes.
- **La phytostabilisation :** les plantes vont limiter l'érosion (action du vent notamment) et immobiliser les polluants dans la zone racinaire (rhizofiltration) en évitant en particulier leur migration vers les eaux de surface et souterraines.
- **La phytodégradation ou phytotransformation :** les plantes vont contribuer directement à la dégradation des polluants en composés moins toxiques au niveau de leurs cellules ou dans l'environnement grâce à certains composés qu'elles produisent comme des enzymes.
- **La phytoextraction :** les plantes sont utilisées pour extraire les polluants du milieu et les concentrer dans des organes destinés à la récolte.
- **La phytovolatilisation :** les plantes sont utilisées pour extraire les polluants du milieu et les transformer en composés volatils qui seront, au final, rejetés dans l'atmosphère.

Lors de la mise en œuvre de ces méthodes, l'implantation des espèces végétales sur les sites contaminés peut être durable, celles-ci auront alors un rôle plutôt phytostabilisant, ou transitoire avec l'utilisation de cultures qui possèdent de fortes capacités d'accumulation et qui seront ensuite récoltées afin d'être traitées (Kramer, 2005).

La phytoremédiation permet de neutraliser la plupart des substances chimiques, qu'elles soient organiques (solvants halogénés, herbicides, explosifs, hydrocarbures aliphatiques, mono-aromatiques, polycycliques (HAP), PCBs), inorganiques (nitrates, phosphates, métaux) ou encore des isotopes radioactifs (^{238}U , ^{137}Cs , ^{90}Sr) (Pilon-Smits, 2005). Pour ce qui est des composés de la famille des chloroéthènes, leur phytoremédiation par différents processus a été observée. Tout d'abord, ils peuvent être phytovolatilisés par plusieurs espèces végétales, le peuplier constituant l'un des meilleurs candidats. En effet, il dispose d'un système racinaire assez étendu, ce qui lui permet d'atteindre des nappes

phréatiques plus profondes, mais il croit également rapidement et consomme de fortes quantités d'eau (Newman *et al.*, 1997; Collins *et al.*, 2002; Ma and Burken, 2003; James *et al.*, 2009). Cependant, d'autres arbres comme le cyprés, le tupelo, le pin (Vroblesky *et al.*, 1998; Nietch *et al.*, 1999) ou des plantes comme la luzerne sont également susceptibles de réaliser ce processus. Les chloroéthènes peuvent également être phytodégradés grâce à l'action d'enzymes synthétisées par certaines plantes. Ainsi, le peuplier, en plus de la volatilisation, peut dégrader le TCE en trichloroéthanol et en acides trichloroacétique et dichloroacétique (Newman *et al.*, 1997). Enfin, les chloroéthènes peuvent être éliminés par phytostimulation grâce à l'association de microorganismes et d'espèces végétales, les premiers possédant des capacités de biodégradation et les seconds stimulant leur développement. C'est le cas de la relation entre le blé et une souche de *Pseudomonas fluorescens* rhizosphérique capable de transformer le TCE par co-métabolisme aérobie (Yee *et al.*, 1998).

Malgré ses nombreux avantages, tels que son faible coût et son aspect écologique (préservation de la faune et de la flore, régénération des milieux, etc), l'utilisation de la phytoremédiation pour la restauration des milieux contaminés par les chloroéthènes reste relativement marginale. En effet, elle présente un certain nombre d'inconvénients qui la rendent difficilement applicable. Le premier concerne la zone couverte par le système racinaire qui varie de quelques dizaines de centimètres pour les herbacées à quelques mètres de profondeur pour certains arbres. Or, en raison de leur accumulation au fond des aquifères et dans des zones difficiles d'accès (Figure 16), les chloroéthènes sont bien souvent hors d'atteinte du système racinaire (Henry *et al.*, 2002; Tyagi *et al.*, 2011). De plus, certaines espèces végétales, comme le peuplier, qui sont capables de phytovolatiliser ces molécules, perdent leurs feuilles en hiver. Une plus faible consommation d'eau durant cette période est alors observée ce qui ne permet pas de décontaminer les zones à traiter de manière efficace en continu. Enfin, puisque certains végétaux peuvent transférer les chloroéthènes jusque dans leurs fruits (pommes, pêches ou tomates) ou les légumes (épinards ou carottes) (Schnabel *et al.*, 1997; Chard *et al.*, 2006; Doucette *et al.*, 2007), leur utilisation en phytoremédiation pourrait mettre en danger la chaîne alimentaire si des animaux ou l'Homme venaient à les consommer.

4.3.2. La bioremédiation microbienne

La bioremédiation microbienne consiste à exploiter les capacités naturelles que possèdent de nombreux microorganismes à dégrader des substances chimiques toxiques en composés inertes, comme l'eau et le gaz carbonique, ou inoffensifs (Vidali, 2001; Megharaj *et al.*, 2011). Le plus souvent, les microorganismes impliqués dans ces processus sont des procaryotes (bactéries ou archées), mais certains champignons ou microalgues ont également été identifiés (Pinedo-Rivilla *et al.*, 2009; Harms *et al.*, 2011). Du fait de leur fort pouvoir d'adaptation, mais aussi de leur immense réservoir génétique, ces microorganismes ont su développer des machineries enzymatiques capables de dégrader la plupart des molécules considérées comme polluantes (Bonaventura and Johnson, 1997; Megharaj *et al.*, 2011). Pour cela, ils peuvent agir seuls ou en *consortium*.

Les microorganismes sont présents dans tous les milieux (sol, aquifère, eaux de surface, sédiments, océans, etc...) où ils peuvent vivre dans des conditions extrêmes (température, salinité, oxygène, etc) (Guerrero and Berlanga, 2006; Tyagi *et al.*, 2011). Les contraintes géologiques et géochimiques ne paraissant pas être un facteur limitant pour l'établissement de certaines populations bactériennes, la bioremédiation microbienne pourrait permettre d'envisager le traitement de n'importe quel site contaminé. Par ailleurs, même si les bactéries anaérobies sont moins faciles à observer et à détecter que celles aérobies, elles présentent un grand intérêt dans la dégradation de molécules récalcitrantes comme certains composés chlorés (Chen, 2004; Smidt and de Vos, 2004; Field and Sierra-Alvarez, 2008; Hiraishi, 2008; Bunge and Lechner, 2009). Cependant, la mise en place d'un traitement de bioremédiation basé sur l'utilisation de microorganismes nécessite le contrôle non seulement de la disponibilité des polluants, de la présence des souches épuratrices mais aussi l'ajustement en permanence d'un certain nombre de conditions environnementales pour maximiser son efficacité : quantité et type de nutriments, présence des différents accepteurs d'électrons (oxygène, nitrate, sulfate...), potentiel redox, pH, température et salinité. Les principales approches utilisées pour atteindre ces objectifs sont l'atténuation naturelle, la biostimulation et la bioaugmentation (Vidali, 2001; Megharaj *et al.*, 2011; Tyagi *et al.*, 2011). Concernant les chloroéthènes, toutes ces stratégies ont fait l'objet de recherches intensives au cours de ces dernières décennies (Semprini, 1995; Grandel and Dahmke, 2004; Löffler and Edwards, 2006; Pant and Pant, 2010).

a) L'atténuation naturelle

L'atténuation naturelle regroupe tous les processus biotiques (biodégradation) et abiotiques (diffusion, advection, sorption, désorption et volatilisation) participant naturellement à la décontamination progressive de l'environnement. Dans la plupart des cas, la biodégradation est le processus le plus efficace (Smets and Pritchard, 2003). Même si cette stratégie ne nécessite pas d'intervention humaine particulière, elle requiert tout de même un suivi permanent afin d'évaluer son évolution. Pour effectuer le suivi du niveau de pollution des nappes phréatiques, de nombreux puits de prélèvements doivent être réalisés de manière à couvrir l'ensemble du panache de contamination. Ils permettent l'analyse régulière de paramètres hydrogéologiques (conductivité, porosité, vitesse...), hydrogéochimiques (redox, oxygène dissous, nitrate, ammonium, fer, manganèse, sulfate, méthane, concentrations en polluants et sous-produits de dégradation, sources de carbone...) et microbiologiques (cultures en microcosmes, analyses de marqueurs moléculaires). Les résultats récoltés servent à implémenter des modèles analytiques nécessaires pour estimer l'impact et la durée des traitements (Grandel and Dahmke, 2004).

Cette stratégie a été appliquée sur des sites contaminés par des chloroéthènes présentant des propriétés géologiques différentes comme des roches fracturées (Lenczewski *et al.*, 2003), des aquifères sableux peu profonds (Nobre and Nobre, 2004) ou des lacs (An *et al.*, 2004). Elle peut aboutir à la minéralisation complète des molécules mères avec production d'ions chlorures, d'eau et de CO₂. Cependant, dans certains cas où la biodégradation anaérobie est incomplète, du DCE et du VC se forment et peuvent s'accumuler notamment lorsque les souches capables de les dégrader ne sont pas présentes naturellement sur le site contaminé (Hendrickson *et al.*, 2002). Ces intermédiaires peuvent ensuite être éliminés par biodégradation aérobie ou de manière abiotique, mais leurs comportements et leurs teneurs doivent être étroitement surveillés. Bien que cette stratégie soit avantageuse au niveau de son coût relativement faible en comparaison des techniques de remédiation physique ou chimique, elle peut cependant nécessiter de nombreuses années de traitement. Elle ne pourra donc être appliquée que pour des contaminations de faible ampleur (Grandel and Dahmke, 2004). Sinon, d'autres techniques visant à améliorer ces processus d'atténuation naturelle soit par modifications de certains paramètres environnementaux (biostimulation), soit par l'apport de souches adéquates (bioaugmentation) peuvent être appliquées (Tyagi *et al.*, 2011).

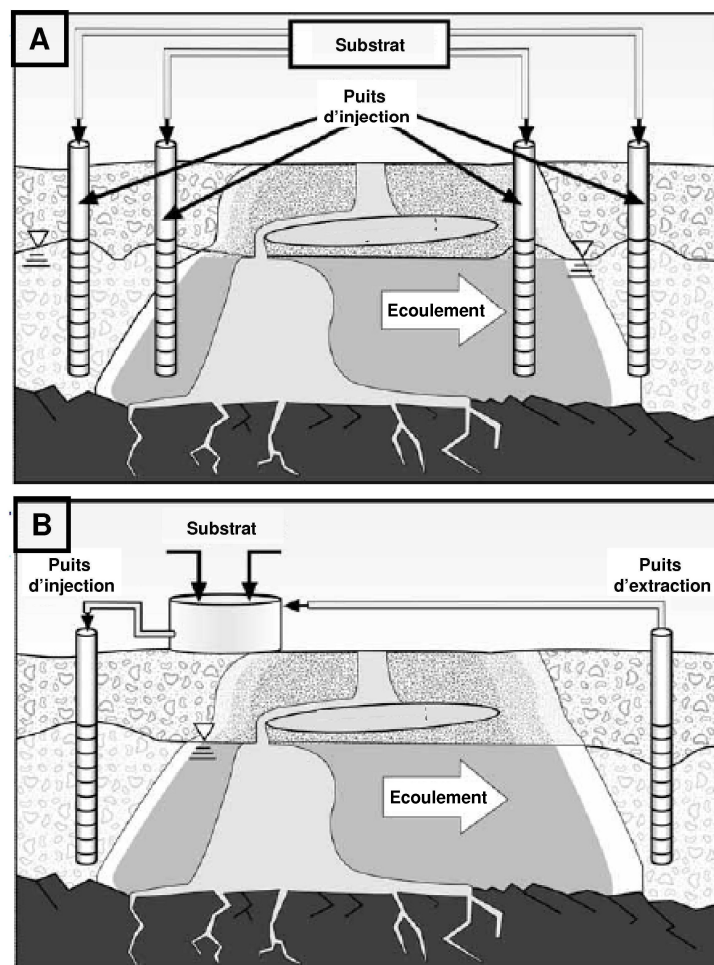


Figure 23 : Principe de l'Enhanced Reductive Dechlorination (ERD) par injection directe de substrat dans l'aquifère (A) ou par la technique de re-circulation (B) (d'après le document « Engineered Approaches to *In Situ* Bioremediation of Chlorinated Solvents: Fundamentals and Field Applications » de l'EPA (2000)).

b) La biostimulation

Le procédé de biostimulation consiste à stimuler l'activité de la microflore indigène en modifiant un ou plusieurs paramètres environnementaux dans le but d'accélérer les processus naturels de biodégradation (Megharaj *et al.*, 2011). Afin de faciliter l'élimination des chloroéthènes, par stimulation des différents processus métaboliques décrits précédemment dans ce manuscrit (paragraphe 2), plusieurs techniques, nécessitant la présence ou non d'oxygène, ont été développées :

- **Le biosparging et le bioventing :** ces techniques de traitement *in situ* allient deux principes de décontamination qui sont la biodégradation et la ventilation. Elles reposent sur l'injection d'air ou de gaz (méthane, propane ou butane) directement dans la zone saturée (biosparging) ou non-saturée (bioventing) afin de stimuler les processus de biodégradation aérobies (co-métabolisme et assimilation aérobies détaillés dans les paragraphes 2.1 et 2.2) (Kim *et al.*, 2008; Davis *et al.*, 2009). L'air injecté mobilise, en les vaporisant, les contaminants dissous dans l'eau ou piégés par capillarité dans les pores. L'apport d'oxygène par l'air favorise ensuite le développement de la biomasse microbienne en provoquant la consommation du carbone des composés chlorés présents dans la phase vaporisée et ainsi leur dégradation. Les autres gaz induisent, quant à eux, la biodégradation des chloroéthènes par des processus co-métaboliques. La mise en œuvre du biosparging et du bioventing est identique à celle décrite pour la technique physique d'air sparging (paragraphe 4.1).
- **L'Enhanced Reductive Dechlorination (ERD) :** cette technique consiste à apporter une source de carbone et/ou de donneurs d'électrons afin de stimuler la biodégradation anaérobie des chloroéthènes par déchloration réductrice (description dans le paragraphe 2.3). Ces apports se font soit par injection directe des substrats dans la nappe phréatique (sous forme de gaz pour l' H_2 et en solution pour les autres composés) (Figure 23A), soit par recirculation (extraction de l'eau de la nappe, ajout du substrat puis réinjection) (Figure 23B), ou encore par la mise en place de barrières bio-réactives qui ressemblent fortement aux barrières réactives chimiques (Figure 20) (EPA, 2000; Kao *et al.*, 2003; Kalin, 2004; Luijten *et al.*, 2004; Löffler and Edwards, 2006; Nishimura *et al.*, 2008).

L'apport direct d' H_2 , principal donneur d'électron utilisé par les microorganismes déhalorespirants, peut être mis en place pour stimuler le processus de déhalorespiration (He *et al.*, 2002; Nishimura *et al.*, 2008). Des sources carbonées sont également fréquemment utilisées sous la forme de molécules organiques simples facilement assimilables par les microorganismes comme l'acétate, le lactate, le formate, le propionate, le butyrate, l'éthanol

Tableau 11 : Réactions de fermentation de molécules carbonées simples conduisant à la production d'H₂ (d'après le document « *In Situ* Bioremediation of Chlorinated Ethene: DNAPL Source Zones » de l'ITRC (2008)).

Substrat	Réaction de fermentation
Ethanol	$C_2H_6O + H_2O \Rightarrow C_2H_3O_2^- + H^+ + 2H_2$ Fermentation de l'éthanol en acétate
Méthanol	$CH_4O + 2H_2O \Rightarrow CO_2 + H_2O + 3H_2$ Fermentation du méthanol
Acétate	$C_2H_3O_2^- + H^+ + 4H_2O \Rightarrow 2CO_2 + 2H_2O + 4H_2$ Fermentation de l'acétate
Butyrate	$C_4H_7O_2^- + 2H_2O \Rightarrow 2C_2H_3O_2^- + H^+ + 2H_2$ Fermentation du butyrate en acétate
Propionate	$C_3H_5O_2^- + 3H_2O \Rightarrow C_2H_3O_2^- + CO_2 + H_2O + 3H_2$ Fermentation du propionate en acétate
Lactate	$C_3H_5O_3^- + 2H_2O \Rightarrow C_2H_3O_2^- + CO_2 + H_2O + 2H_2$ Fermentation du lactate en acétate

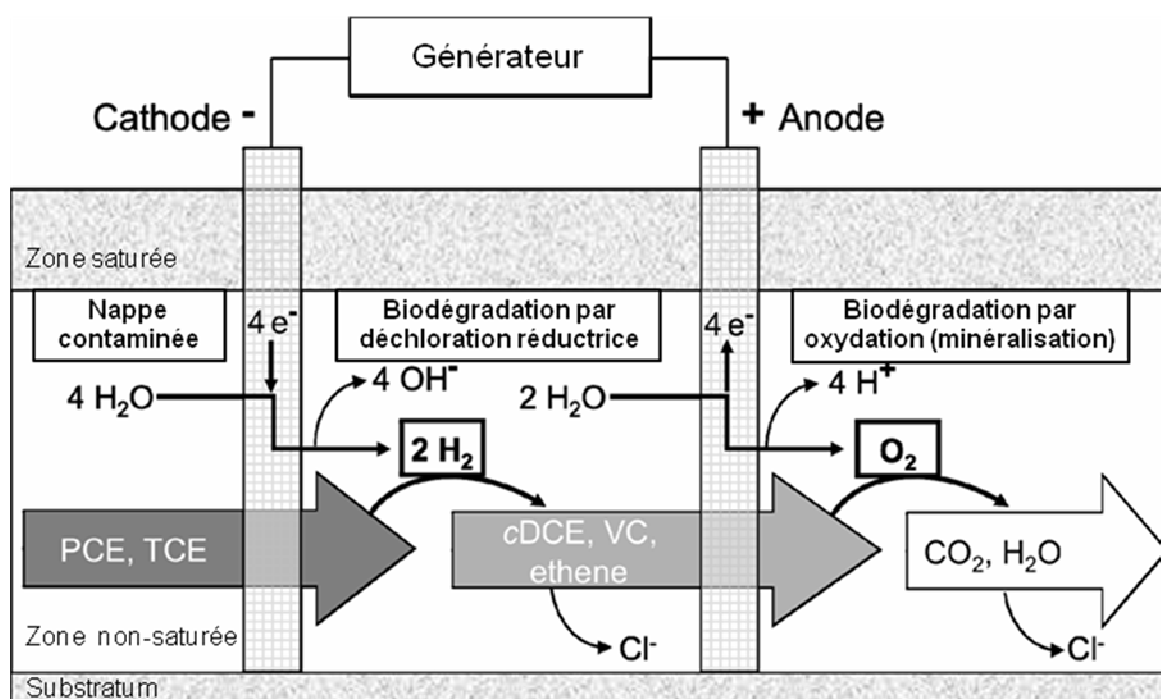


Figure 24 : Dispositif d'électro-stimulation séquentielle anaérobie/aérobie mis en place sur un site contaminé par du PCE (d'après L'henner *et al.*, 2011).

ou le méthanol (Luijten *et al.*, 2004; Azizian *et al.*, 2010). Elles peuvent être soit directement utilisées comme donneurs d'électrons par certains microorganismes déhalorespirants facultatifs (Tableau 5), soit fermentées par d'autres microorganismes, provoquant ainsi la consommation de l'oxygène résiduel avec production de sous-produits (comme l'H₂) qui serviront de donneurs d'électrons pour les microorganismes déhalorespirants obligatoires (Tableau 11). Enfin, une dernière solution consiste à injecter des molécules plus complexes comme des huiles végétales (Long and Borden, 2006), des boues, de la mélasse (Kao *et al.*, 2003), de la tourbe (Kao *et al.*, 2001) ou du tétrabutoxysilane (Yu and Semprini, 2009) dont la consommation produira une libération d'H₂ plus lente et plus étendue dans le temps.

- **Le traitement séquentiel anaérobie/aérobie :** la biodégradation du PCE et du TCE par déchloration réductrice peut conduire, en l'absence de certaines souches de *Dehalococcoides* ou dans des conditions défavorables, à l'accumulation de DCE et de VC dans l'aquifère (Hendrickson *et al.*, 2002). Or, ces produits de dégradation peuvent être pris en charge de manière très efficace au cours de processus aérobies (que ce soit par la voie co-métabolique ou la voie d'assimilation) (voir paragraphes 2.1 et 2.2) (Mattes *et al.*, 2010). C'est pourquoi une technique consistant à stimuler séquentiellement la dégradation des chloroéthènes par les voies anaérobies et aérobies a été développée (Beeman and Bleckmann, 2002; Tiehm and Schmidt, 2011). Dans l'étude de Beeman et Bleckmann (2002), les conditions réductrices nécessaires au processus de déchloration anaérobie ont été mises en place dans la zone source de contamination grâce à l'approche ERD (technique de recirculation), suivie du pompage de l'eau du panache, de son oxygénation et de sa réinjection dans des puits situés en aval hydraulique dans le but de stimuler les processus aérobies. Le site contaminé utilisé pour cette étude a ainsi pu être entièrement réhabilité par cette stratégie, alors qu'aucun des précédents traitements de biostimulation (aérobies ou anaérobies) ne l'avait permis (Beeman and Bleckmann, 2002).

- **L'électro-stimulation :** l'application d'un courant électrique dans l'aquifère par l'intermédiaire d'électrodes peut permettre de générer des électrons au niveau de la cathode ou de l'oxygène au niveau de l'anode. Les électrons peuvent être captés par les microorganismes capables de transformer les chloroéthènes par le processus de déchloration réductrice (Aulenta *et al.*, 2009). A l'inverse, l'oxygène libéré par l'anode peut stimuler les processus de dégradation aérobies (co-métabolisme et assimilation) (Guiot *et al.*, 2008). Les deux réactions peuvent être utilisées séparément. Cependant, elles ont récemment été mises en place simultanément dans une stratégie d'électro-stimulation séquentielle anaérobie/aérobie pour la décontamination réussie d'un site pollué par du PCE (Figure 24) (Lohner *et al.*, 2011).

c) La bioaugmentation

Lorsque les approches d'atténuation naturelle ou de biostimulation sont mises en échec, notamment du fait de l'absence ou de la trop faible abondance des souches microbiennes capables de dégrader les polluants, il est possible d'avoir recours à une autre alternative : la bioaugmentation. Ce procédé consiste à ajouter des microorganismes dans la zone polluée afin d'augmenter la biodégradation des contaminants. Les microorganismes ajoutés peuvent être indigènes ou alors étrangers au milieu. Dans le premier cas, ils sont extraits du milieu, acclimatés au contaminant, cultivés *ex situ* puis réinjectés. Le procédé de bioaugmentation est envisagé lorsque la biostimulation de la flore indigène ne parvient pas à augmenter les vitesses de biodégradation (Vogel, 1996).

L'ensemencement des sites contaminés peut s'effectuer à l'aide d'un inoculum composé d'une souche seule, comme par exemple la souche *Mycobacterium* sp. TRW-2 capable de dégrader le *cis*-DCE, le VC et l'éthylène en conditions aérobies (Fathepure *et al.*, 2005). Cependant, la biodégradation anaérobie par déchloration réductrice, qui est très efficace pour l'élimination des chloroéthènes les plus substitués (PCE et TCE), est souvent une entreprise collective menée par des individus appartenant à différentes espèces d'un même genre, ou à différents genres bactériens. Par exemple, la coopération a pu être montrée entre des membres de *Desulfitobacterium* et *Dehalococcoides* (Yang *et al.*, 2005; Bunge *et al.*, 2007; Rouzeau-Szynalski *et al.*, 2011), *Dehalobacter* et *Dehalococcoides* (Daprato *et al.*, 2007), *Geobacter* et *Dehalococcoides* (Duhamel and Edwards, 2007; Amos *et al.*, 2009), *Sulfurospirillum* et *Dehalococcoides* (Maillard *et al.*, 2011) ou encore plusieurs souches de *Dehalococcoides* (Duhamel *et al.*, 2002; Holmes *et al.*, 2006). Au sein des nombreux *consortia* isolés montrant de fortes capacités de dégradation, figurent presque systématiquement des souches affiliées au genre *Dehalococcoides*, seuls microorganismes capables de réaliser les dernières étapes de la déchloration réductrice (Maymo-Gatell *et al.*, 1995; DeWeerd *et al.*, 1998; Richardson *et al.*, 2002; He *et al.*, 2003; Duhamel *et al.*, 2004; Vainberg *et al.*, 2009; Cichocka *et al.*, 2010; Ziv-El *et al.*, 2011). Les autres partenaires de ces *consortia* peuvent alors soit participer à la déchloration réductrice, soit intervenir dans la fermentation des substrats pour former de l'H₂ ou dans la production de co-facteurs essentiels (voir paragraphe 2.4.1). Ainsi, ces *consortia* sont largement utilisés en bioaugmentation (Ellis *et al.*, 2000; Major *et al.*, 2002; Scheutz *et al.*, 2008; Scheutz *et al.*, 2010).

Lorsque la bioaugmentation vise à augmenter le processus de déchloration réductrice anaérobie, elle est généralement couplée à des techniques de biostimulation comme l'ERD.

L'inoculum est alors ajouté en même temps que le donneur d'électron (Scheutz *et al.*, 2008; Scheutz *et al.*, 2010). Pour cela, les *consortia* sont dans un premier temps produits en bioréacteurs ou directement extraits d'environnements au sein desquels l'activité de déchloration réductrice est avérée. Puis, ils sont dispensés dans des récipients scellés afin de maintenir l'anaérobiose durant l'acheminement sur le site contaminé et enfin il sont distribués dans les puits d'injection (Figure 23A) (Löffler and Edwards, 2006). En raison de leur forte sensibilité à l'oxygène, l'exposition des souches de *Dehalococcoides* à l'air ambiant doit être évitée durant ce processus.

4.4. Futures directions

La restauration des aquifères contaminés par les chloroéthènes est devenue un des enjeux économiques et écologiques majeur de notre société. Les études pluridisciplinaires menées ces dernières décennies, mettant en commun les compétences de biologistes, de physiciens, de chimistes et d'ingénieurs, ont permis de proposer des techniques opérationnelles pour une gestion intégrée de la pollution des milieux, mais aussi des procédés de dépollution très efficaces (Henry *et al.*, 2002; Khan *et al.*, 2004; Scullion, 2006). Les exemples de succès les plus récents montrent que les associations entre plusieurs techniques peuvent donner des résultats très intéressants en terme d'efficacité de traitement (Christ *et al.*, 2005). Ces associations peuvent impliquer des techniques physiques et chimiques (Li and Hanlie, 2008), des techniques physiques et biologiques (Ramsburg *et al.*, 2004; Goltz *et al.*, 2005; Friis *et al.*, 2007; Lohner *et al.*, 2011) ou encore des techniques provenant de l'ensemble de ces disciplines (Tsai *et al.*, 2009).

L'utilisation de la biologie synthétique, avec la création de nouvelles souches bactériennes capables de dégrader efficacement de multiples polluants par génie génétique (OGM), pourrait également permettre, sous condition d'autorisation par les lois de bioéthique, de proposer dans un futur proche des solutions de bioaugmentation particulièrement bien adaptées (Furukawa, 2003, 2006). Cependant, si la création de microorganismes porteurs des gènes cataboliques impliqués dans les voies de dégradation aérobies semble assez bien maîtrisée, le transfert de voies métaboliques anaérobies dans une cellule hôte appropriée n'a pas encore été décrit (Ramos *et al.*, 2011). Enfin, depuis la création d'une souche bactérienne contrôlée par un génome d'origine totalement synthétique, *Mycoplasma mycoides* JCVI-syn1.0 (Gibson *et al.*, 2010), le concept de cellule artificielle est sur le point de devenir une réalité (Noireaux *et al.*, 2011). La création de microorganismes « à façon » pour des

applications en bioremédiation serait une des nombreuses opportunités de profiter de cette avancée technologique majeure.

Dans le contexte socio-économique actuel, les approches de bioremédiation, et en particulier celles utilisant des microorganismes non manipulés génétiquement, offrent la possibilité de réhabiliter la plupart des sites pollués à faible coût tout en limitant les impacts sur les écosystèmes. De plus, elles sont plutôt bien accueillies par l'opinion publique (approche « écologique »). Cependant, si les processus de biodégradation utilisés sont bien maîtrisés en conditions de laboratoire, leurs applications sur le terrain par les professionnels de la dépollution restent encore aujourd'hui relativement empiriques (boîte noire). En effet, la caractérisation de la microflore indigène présente *in situ*, l'identification et le suivi des populations épuratrices ou encore les mesures d'impact des différents traitements sur la microflore impliquée dans le « biosystème » de dépollution, ne sont que très rarement évalués, faute d'outils adaptés. C'est pourquoi le développement d'approches moléculaires performantes permettant de caractériser l'ensemble des microorganismes impliqués dans le processus d'épuration ainsi que de leurs voies métaboliques reste un enjeu majeur pour l'établissement de stratégies de bioremédiation efficaces (Lovley, 2003; Xu, 2006; Maphosa *et al.*, 2010).

5. Les apports des outils « haut débit » pour le développement de stratégies de bioremédiation

Pour l'élaboration de stratégies de bioremédiation microbienne efficaces, il est indispensable de connaître pour chaque site pollué (i) l'identité des individus composant la microflore, (ii) leurs capacités métaboliques mais aussi (iii) leurs réponses aux changements de conditions environnementales. Or, encore aujourd'hui, ces données sont extrêmement difficiles à acquérir. En effet, la tâche est immense puisqu'on estime qu'il existerait sur Terre environ 5×10^{30} cellules bactériennes (Whitman *et al.*, 1998), ce qui représente une biomasse équivalente à celle des végétaux, et qu'il y aurait entre plusieurs milliers et plusieurs dizaines de milliers d'espèces bactériennes différentes par gramme de sol (Torsvik *et al.*, 1990; Curtis *et al.*, 2002; Gans *et al.*, 2005). Par conséquent, notre connaissance de la diversité microbienne est en partie limitée par notre difficulté à l'étudier.

Jusque dans les années 1990, l'une des principales contraintes de la microbiologie environnementale résidait dans l'obligation de passer par la mise en culture des microorganismes pour réaliser leur identification ou l'étude de leurs capacités métaboliques. Or, la proportion, estimée aujourd'hui à environ 1%, de bactéries cultivées issues de milieux naturels est très faible (Torsvik and Øvreås, 2002). Depuis une décennie, le développement de nouvelles techniques telles que la séparation des microorganismes soit par cytométrie en flux soit *via* l'utilisation de la micromanipulation, des techniques de filtration ou de centrifugation, la co-culture avec des souches « helper », l'élaboration de chambres de diffusion ou encore l'utilisation de milieux simulant des environnements naturels (Vartoukian *et al.*, 2010), ont permis d'augmenter considérablement (plusieurs centaines de fois) la possibilité d'isoler de nouveaux microorganismes. Cependant, malgré ces progrès, elles ne donnent toujours accès qu'à une partie infime de l'immense diversité microbienne. Par ailleurs, ces approches sont encore souvent inadaptées pour l'isolement de microorganismes provenant d'environnements complexes perturbés. En effet, ceux-ci possèdent généralement un mode de vie en *consortia* et nécessitent, pour se développer, la présence de molécules apportées par d'autres membres de la communauté microbienne. Ainsi, si l'isolement de certaines espèces peut être réalisé avec succès, leur maintien et leur transfert dans du milieu frais sont difficiles, voire impossibles, du fait de la difficulté d'identifier le(s) facteur(s) manquant(s) indispensable(s) à leur croissance. Les chercheurs ont été confrontés à ce problème pour certaines souches affiliées au genre

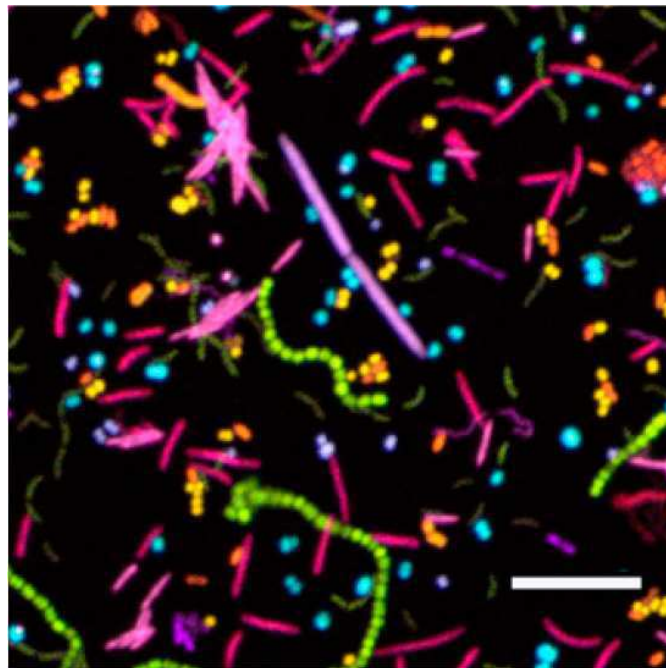


Figure 25 : Caractérisation du microbiote oral humain par la technique FISH grâce à l'utilisation d'un mélange de sondes ciblant 15 *taxa* (genres ou familles) classiquement retrouvés dans la cavité buccale (d'après Valm *et al.*, 2011). Chaque sonde est synthétisée en deux exemplaires, chacun marqué avec un fluorochrome différent parmi la liste suivante : Alexa Fluor 488, Alexa Fluor 514, Alexa Fluor 555, Alexa Fluor 594, Alexa Fluor 647, et Rhodamine Red-X. De cette façon chaque taxon est ciblé par une combinaison unique de deux fluorochromes. Liste des 15 *taxa* ciblés : *Streptococcus*, *Fusobacterium*, *Leptotrichia*, *Veillonella*, *Neisseriaceae*, *Porphyromonas*, *Capnocytophaga*, *Pasteurellaceae*, *Gemella*, *Selenomonas*, *Prevotella*, *Actinomyces*, *Rothia*, *Campylobacter* et *Treponema*.

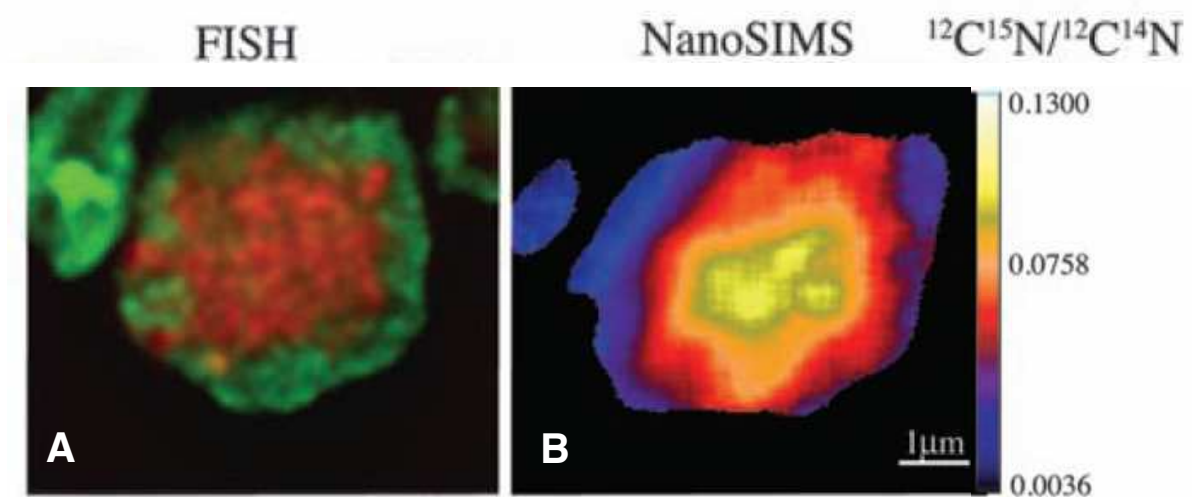


Figure 26 : Analyse FISH-nanoSIMS de l'incorporation d'azote par le *consortia* ANME-2/DSS incubé avec du $^{15}\text{N}_2$ (d'après Dekas *et al.*, 2009). Le *consortium* est composé de deux populations microbiennes : des archées capables de réaliser l'oxydation anaérobie du méthane (ANME-2) et des bactéries sulfato-réductrices appartenant au groupe *Desulfosarcina/Desulfococcus* (DSS). A : image FISH obtenue après l'hybridation de sondes ciblant le groupe ANME-2 (en rouge) et le groupe DSS (en vert). B : distribution du rapport $^{12}\text{C}^{15}\text{N}/^{12}\text{C}^{14}\text{N}$. Cette analyse révèle que c'est la population ANME-2 qui est responsable de la fixation du N_2 au sein du *consortia*.

Dehalococcoides comme *D. ethenogenes* 195 (Maymo-Gatell *et al.*, 1997), ce qui complique l'étude *in vitro* de leurs capacités métaboliques.

Afin d'acquérir des données complémentaires aux approches culturales et de visualiser directement l'abondance, la répartition et les interactions de microorganismes d'intérêt *in situ*, des approches d'observation basées sur l'utilisation de sondes fluorescentes ciblant une séquence d'acides nucléiques particulière ont alors vu le jour (Figure 25). Les plus couramment utilisées aujourd'hui sont le FISH (Fluorescent *In Situ* Hybridization) et le CARD-FISH (DeLong *et al.*, 1989; Amann *et al.*, 1990; Schonhuber *et al.*, 1997; Bottari *et al.*, 2006; Valm *et al.*, 2011). Elles ont d'ailleurs été appliquées pour la détection des *Dehalococcoides* au sein d'échantillons environnementaux grâce à la détermination de sondes spécifiques de ce groupe (Yang and Zeyer, 2003). Cependant, même si ces techniques génèrent des données quantitatives intéressantes, elles ne fournissent pas d'informations suffisantes sur les capacités d'adaptations de ces microorganismes. Pour combler ce manque, plusieurs stratégies couplant le FISH à des techniques isotopiques ont été développées. Ainsi, la combinaison de la microautoradiographie (MAR) et du FISH (Lee *et al.*, 1999), appelée MAR-FISH, donne accès à la fois à l'identité et à l'activité métabolique d'un microorganisme présent au sein d'une communauté complexe. Plus récemment, le couplage du NanoSIMS (spectromètre de masse à ionisation secondaire à l'échelle nanométrique) et du FISH a donné naissance à la technique appelée SIMSISH (ou FISH-NanoSIMS) (Figure 26). En plus d'une meilleure résolution pour l'observation des communautés microbiennes complexes, cette stratégie autorise également la détection de tous les éléments du tableau périodique (Li *et al.*, 2008).

Malgré l'évolution constante des approches décrites précédemment, permettant maintenant de caractériser plus finement les microorganismes d'intérêt au sein de communautés microbiennes complexes, leur application reste limitée non seulement par la difficulté de leur mise en œuvre mais surtout par leur faible débit. En effet, elles n'autorisent pas l'étude simultanée d'un grand nombre d'échantillons et ne fournissent d'informations que pour un nombre restreint de microorganismes, ce qui est insuffisant pour comprendre le fonctionnement d'un écosystème dans sa globalité.

5.1. Révolution des approches moléculaires et émergence des outils haut débit

L'essor spectaculaire des approches moléculaires réalisé ces vingt dernières années, permet aujourd'hui d'aborder les problématiques d'écologie microbienne simplement au travers de l'analyse des molécules d'ADN et/ou d'ARN sans passer par une étape de culture ou d'observation sous microscope (Amann *et al.*, 1995; Pace, 1997). Avec ces approches, l'étude des communautés microbiennes repose le plus souvent sur l'utilisation de biomarqueurs qui sont capables de donner des renseignements sur l'identité (biomarqueurs phylogénétiques) ou le rôle fonctionnel (biomarqueurs fonctionnels) d'un grand nombre de microorganismes. Le principal biomarqueur phylogénétique utilisé en microbiologie est le gène codant pour la petite sous unité de l'ARN ribosomique (ARNr 16S chez les procaryotes et ARNr 18S chez les eucaryotes) (Woese *et al.*, 1990). Certaines signatures de ce biomarqueur peuvent, en outre, être caractéristiques de groupes fonctionnels. Cependant, les biomarqueurs fonctionnels sont généralement des gènes codant pour des enzymes impliquées dans des métabolismes d'intérêt. Ainsi aujourd'hui, la caractérisation fonctionnelle des souches affiliées au genre *Dehalococcoides* se fait principalement par l'intermédiaire des gènes codant pour les déhalogénases réductrices (RdhA) qui, comme cela est décrit dans les paragraphes 2.4 et 3.1, sont caractéristiques de la déchloration réductrice et sont donc d'excellents biomarqueurs de ce processus métabolique (Tas *et al.*, 2010). Notons simplement qu'à l'image des évolutions décrites dans le paragraphe précédent pour les techniques d'observation, certaines approches moléculaires comme la SIP (Stable Isotope Probing) utilisent l'incorporation d'isotopes stables, provenant de substrats marqués, dans les molécules d'ADN pour permettre de relier l'identification des microorganismes à leur fonction métabolique (Lueders *et al.*, 2004).

5.1.1. Les méthodes basées sur l'amplification PCR

La technique de réaction de polymérisation en chaîne (PCR) (Saiki *et al.*, 1985), repose sur l'amplification exponentielle et spécifique d'une portion d'ADN ciblée grâce à un couple d'amorces. Son utilisation ainsi que celle de ses nombreuses variantes (RT-PCR, RACE-PCR, PCR multiplex, qPCR, qRT-PCR, etc...) ont permis une avancée majeure dans l'étude des communautés microbiennes présentes dans les milieux environnementaux. Il est ainsi possible de cibler spécifiquement n'importe quelle population ou groupe de

microorganismes pour lesquels des informations de séquences sont disponibles. Pour les *Dehalococcoides*, par exemple, Hendrickson *et al.* (2002) ont déterminé un couple d'amorces spécifique du gène ADNr 16S, ce qui permettait de les mettre en évidence quel que soit l'écosystème étudié. Depuis ce travail, et grâce à l'enrichissement des bases de données de séquences, le nombre de couples d'amorces disponibles pour l'étude des microorganismes capables de dégrader les chloroéthènes a largement augmenté, que ce soit pour cibler leur ADNr 16S ou leurs gènes fonctionnels (He *et al.*, 2003; Smits *et al.*, 2004; Johnson *et al.*, 2005; Holmes *et al.*, 2006; Rahm *et al.*, 2006; Ritalahti *et al.*, 2006; Fung *et al.*, 2007; Cupples, 2008; Nishimura *et al.*, 2008; Yan *et al.*, 2009b; Jin and Mattes, 2010; Miyata *et al.*, 2010). Ainsi, grâce à l'utilisation en parallèle d'amorces ciblant les biomarqueurs phylogénétiques et fonctionnels, Scheutz *et al.* (2008) ont suivi l'abondance du groupe des *Dehalococcoides* (biomarqueur ADNr 16S) et plus particulièrement des populations possédant le gène *vcrA* par une approche PCR quantitative (qPCR) réalisée sur 45 échantillons de nappe phréatique collectés sur un site pollué tout au long de son traitement de bioremédiation par ERD. Au cours de ce travail, ils ont montré une forte corrélation entre la production d'éthylène sur le site et la croissance des *Dehalococcoides* portant le gène *vcrA*. L'apparition de nouvelles stratégies de PCR, utilisant des amorces dégénérées afin de cibler des combinaisons de séquences non répertoriées dans les bases de données, offre, par ailleurs, la possibilité de caractériser des séquences encore jamais observées (Rose *et al.*, 1998; Rose *et al.*, 2003). Grâce à cette stratégie, de nouvelles déhalogénases réductrices putatives ont été isolées non seulement à partir de souches ou *consortia* capables de déhalorespiration (von Wintzingerode *et al.*, 2001; Regard *et al.*, 2004; Futamata *et al.*, 2009), mais également à partir d'échantillons d'ADN issus de sites contaminés (Takeuchi *et al.*, 2011) ou de sédiments marins non pollués (Futagami *et al.*, 2009).

Si les méthodes basées sur la PCR, couplées ou non à des techniques d'analyse d'amplicons comme les techniques d'empreintes génétiques (DGGE, TGGE, ARDRA, T-RFLP, RISA ou encore SSCP), apportent rapidement des informations pertinentes sur la structure et/ou la fonction des communautés microbiennes, celles-ci restent encore incomplètes. En effet, seul un nombre limité de biomarqueurs peut être étudié à la fois, ce qui permet difficilement d'appréhender le fonctionnement d'un écosystème dans son ensemble. De plus, elles sont soumises à de nombreux biais inhérents à la technique de PCR, l'un des principaux étant la spécificité des amorces, ce qui peut conduire à une surreprésentation de certains fragments et altérer la vision des communautés microbiennes présentes dans l'environnement (Suzuki and Giovannoni, 1996; von Wintzingerode *et al.*, 1997; Polz and

Tableau 12 : Comparaison des points forts et des principales limites des différentes plateformes de séquençage.
ND : non déterminé. SMRT : Single Molecule Real Time Technology

	1 ^{ère} génération	2 ^{ème} génération				3 ^{ème} génération			
Plateforme	Séquenceurs ABI (Life Technologies)	454 GS FLX (Roche)	HiSeq 2000 System (Illumina)	SOLiD (Life Technologies)	tSMS (Helicos BioSciences)	PacBio RS (Pacific Biosciences)	Ion torrent (Life Technologies)	Oxford Nanopore Technology	FRET (Life Technologies)
Amplification	PCR avec des ddNTPs	PCR en émulsion	PCR en « ponts »	PCR en émulsion	Séquençage de molécules individuelles	Séquençage de molécules individuelles			
Méthode de séquençage	Sanger	Pyroséquençage	Séquençage par synthèse	Séquençage par ligation	Séquençage par synthèse	SMRT	Semi-conducteur	Nanopore	Transfert d'énergie de résonance
Longueur des lectures (bases)	1500	450 (Titanium) / 700 (FLX)	36-200	50-75	25-55	>1000 (jusqu'à 10000)	200 (400 en 2012)	>100000	>100000
Débit (Gb par run)	0,001	0,45 (Titanium) / 0,7 (FLX)	95-600	10-20	21-35	100	25	ND	ND
Coût par run (approximatif)	8000-10000 \$	8500 \$	9000 \$	17000 \$	ND	Inférieurs aux technologies de 2 ^{ème} génération			
Temps de run	Plusieurs semaines	10 h (Titanium) / 23 h (FLX)	1,5-11 j	6 j	< 1j	30 min	2h	15 min - 1 j	ND
Disponible	oui	oui	oui	oui	oui	oui (avant fin 2011 en Europe)	oui	non	non

Cavanaugh, 1998; Lueders and Friedrich, 2003; Acinas *et al.*, 2005; Engelbrektson *et al.*, 2010).

5.1.2. Les approches haut débit dites de méta « omiques »

Encore aujourd'hui, le rôle exact de la plupart des populations microbiennes au sein de leur milieu reste méconnu. Grâce aux approches dites méta « omiques », comme la métagénomique, la métatranscriptomique ou encore la métaprotéomique, il est désormais possible d'avoir une vision plus intégrative de l'ensemble des événements se déroulant dans un écosystème. Ceci est en grande partie possible grâce au développement de nouveaux outils parmi lesquels figurent les technologies de séquençage. En dix ans, l'évolution des techniques de séquençage de l'ADN a été spectaculaire (comprenant les machines, l'informatique, les outils de traitement et de stockage de données, etc....) ce qui a rendu possible l'exploration des métagénomes, c'est-à-dire de l'ensemble des potentialités génétiques contenu dans un échantillon environnemental (Handelsman *et al.*, 1998).

Bien que la première génération de méthode de séquençage, basée sur la technique de Sanger, soit maintenant moins utilisée, principalement du fait de son coût élevé et de son faible débit (Tableau 12), elle a permis la réalisation de projets d'envergure comme le séquençage du métagénome de la mer des Sargasses (Venter *et al.*, 2004), celui de la surface des océans (partie du projet intitulé « the Sorcerer II Global Ocean Sampling Expedition ») (Rusch *et al.*, 2007) ou encore plus récemment, celui des communautés microbiennes d'une nappe phréatique contaminée à la fois par des métaux lourds, de l'acide nitrique et des solvants organiques (Hemme *et al.*, 2010). La technique de Sanger a laissé la place aux techniques de seconde génération ou Next Generation Sequencing (NGS) (Ansorge, 2009; Metzker, 2010). Ces nouvelles techniques s'affranchissent de l'étape de clonage des fragments d'ADN et permettent de réduire fortement les coûts et les temps d'acquisition des données (Tableau 12). Grâce à ces évolutions, de nombreux projets de métagénomique ont été initiés avec aujourd'hui 337 programmes référencés dans la base de données Genome OnLine Database (GOLD : www.genomesonline.org) dont 68% concernent des métagénomes environnementaux. Ceci a par ailleurs ouvert la porte à des projets extrêmement ambitieux avec, par exemple, une étude récente faisant état de l'utilisation du pyroséquençage pour comparer la diversité fonctionnelle de 45 biomes et de 42 viromes différents au travers de plus de 15 millions de séquences (Dinsdale *et al.*, 2008). Cette approche de métagénomique comparative a conforté l'idée qu'une grande partie de la diversité fonctionnelle était

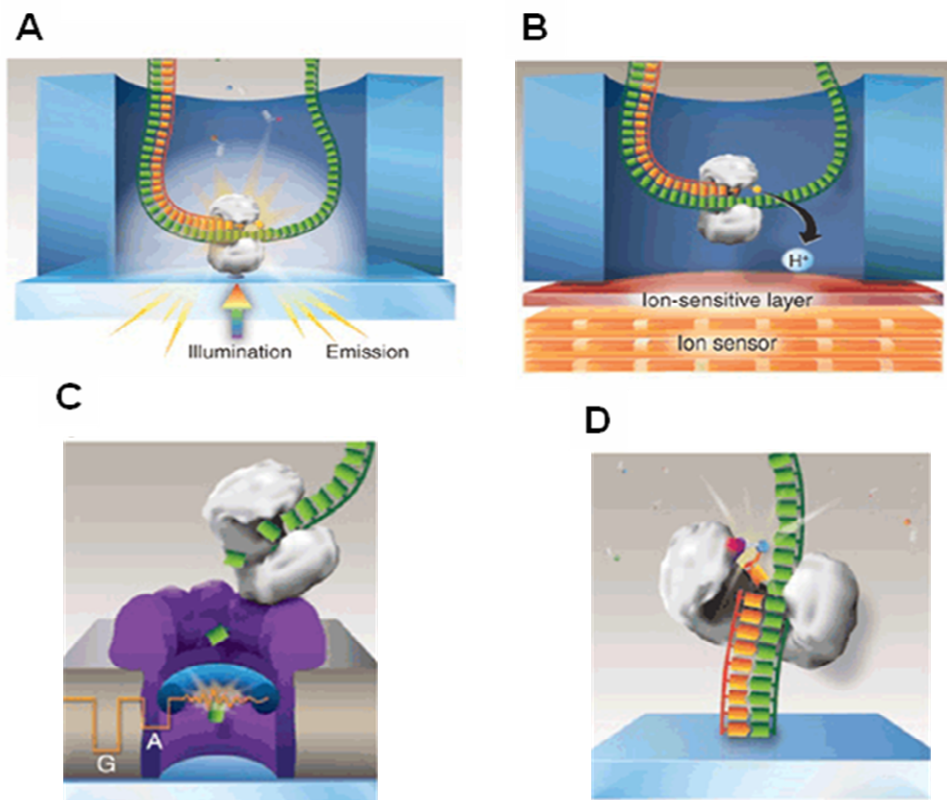


Figure 27 : Principe des technologies de séquençage de troisième génération (d'après Munroe et Harris, 2010).
 A : technologie SMRT (Single Molecule Real Time Technology) utilisée dans le système PacBio RS de Pacific Biosciences. B : stratégie utilisée par le système Ion Torrent de Life Technologies avec l'emploi d'un semi-conducteur. C : technologie Nanopore. D : technologie FRET (Fluorescence Resonance Energy Transfer).

commune à tous les biotopes, tout en montrant que chacun d'entre eux se différencie par la présence de populations microbiennes qui étaient spécifiques de leurs caractéristiques biogéochimiques. Le consortium TerraGenome propose d'ailleurs de mettre en commun les efforts de nombreux partenaires pour créer un métagénome de référence du sol, au travers de l'analyse et de l'annotation de plus de 100 Gb de données de pyroséquençage (Vogel *et al.*, 2009). Ce projet est en cours de réalisation à l'heure actuelle.

Aujourd'hui, le problème majeur des NGS réside dans la taille des fragments séquencés qui est relativement courte, ce qui complique fortement l'assemblage et la reconstruction de génomes complets (Morales and Holben, 2010). De plus, malgré l'immense quantité de données générées, ceci ne suffit pas à accéder aux génomes de l'ensemble des populations microbiennes présentes dans l'environnement. En effet, Quince *et al.* (2008) ont estimé que pour certains écosystèmes particulièrement diversifiés, l'effort de séquençage actuel devrait être multiplié par plus de 10 000 pour couvrir 90% de la diversité qu'ils renferment, ce qui est encore inenvisageable que ce soit sur le plan technique ou économique (Quince *et al.*, 2008). Depuis peu, une troisième génération de technologies de séquençage fait donc son apparition exploitant au maximum les progrès des nanotechnologies afin de simplifier de façon drastique les techniques de séquençage (Tableau 12) (Blow, 2008; Munroe and Harris, 2010; Pareek *et al.*, 2011). Pour cela, les nouveaux procédés se basent sur l'utilisation d'enzymes immobilisées individuellement sur un support et capables de séquencer une seule molécule d'ADN à la fois (Figure 27). Deux d'entre eux sont déjà disponibles sur le marché. Il s'agit du système PacBio RS commercialisé par Pacific Biosciences basé sur la technologie dite SMRT pour « Single Molecule Real Time Technology » (Eid *et al.*, 2009; Korlach *et al.*, 2010; McCarthy, 2010), et du système Ion Torrent de Life Technologies (Rothberg *et al.*, 2011), les deux techniques utilisant des réseaux de nanopuits dans lesquels une polymérase réalise les réactions de séquençage. Le système PacBio RS est constitué d'une structure nanophotonique, appelée "zero-mode wave guide", composée de 80 000 puits de 70 nm de diamètre et 100 nm de long dans une couche d'aluminium déposée sur un verre ce qui assure une détection efficace des signaux fluorescents émis lors de l'incorporation des nucléotides par la polymérase et permet le séquençage en parallèle de nombreux brins (Figure 27A). Le système Ion torrent, quant à lui, n'utilise pas de molécules fluorescentes dans son système de détection mais un semi-conducteur (Figure 27B). En effet, l'incorporation d'un nucléotide par la polymérase va produire un ion H⁺, ce qui va créer transitoirement une différence de potentiel détectable au fond du puits. Pour ce procédé, les nucléotides sont donc apportés les uns après les autres.

Cette troisième génération améliore considérablement le rendement de séquençage et permet ainsi d'obtenir des séquences de taille supérieure (400 bases pour Ion Torrent et >1000 bases pour PacBio RS) à une vitesse proche de 10 bases par seconde. Cependant, les chercheurs veulent aller encore plus loin avec le séquençage de molécules de très grande taille (>100Kb) grâce au développement de deux autres systèmes toujours en cours de finalisation. Pour cela, la technologie Nanopore propose un procédé dans lequel une exonucléase est fixée sur un nanopore d' α -hémolysine (Figure 27C) (Clarke *et al.*, 2009; Timp *et al.*, 2010) alors que la technologie FRET (Fluorescence Resonance Energy Transfer) se sert d'une polymérase modifiée avec un quantum dot pour réduire les temps d'excitation du laser très dommageables pour l'enzyme (Figure 27D) (Munroe and Harris, 2010). Ces nouvelles approches faciliteront sans nul doute les reconstructions de génomes complets à partir d'échantillons métagénomiques. L'exactitude des séquences générées par ces techniques est également un paramètre important à considérer. De même, les étapes d'annotation des séquences demeurent essentielles. Notons que la difficulté d'identifier de nouvelles fonctions ou de nouveaux gènes codant pour des protéines à activités connues peut être contournée par les cribles fonctionnels des banques métagénomiques (Rondon *et al.*, 2000; Tao *et al.*, 2011). Cette dernière approche est actuellement limitée par le nombre restreint de systèmes hétérologues assurant l'expression des gènes de métagénomes (Uchiyama and Miyazaki, 2009).

Les autres approches méta « omiques » avec la métatranscriptomique (Leininger *et al.*, 2006) et la métaprotéomique (Wilmes and Bond, 2004) ont également bénéficié d'évolutions méthodologiques et technologiques intéressantes mais à des niveaux moindres, ce qui a ralenti leur essor comparé à la métagénomique. Cependant, de plus en plus d'études décrivent une utilisation combinée de plusieurs approches méta « omiques » pour répondre aux problématiques environnementales. Ainsi, Bertin et ses collaborateurs, en couplant des approches de métagénomique et de métaprotéomique, ont pu reconstruire les génomes de sept membres de la communauté microbienne d'un écosystème fortement contaminé par de l'arsenic et mettre en évidence la production de plus de 500 protéines impliquées dans divers processus métaboliques (oxydation du fer, du sulfate et de l'arsenic) mais aussi dans de nombreuses fonctions liées aux mécanismes d'interactions syntrophiques (Bertin *et al.*, 2011).

Même si ces technologies émergentes sont très prometteuses pour des applications en écologie microbienne, et plus particulièrement pour l'étude de milieux perturbés, leur mise en place à l'échelle d'un laboratoire est freinée non seulement par le coût d'équipement en machines performantes (et en perpétuelle évolution) mais surtout par la quantité de données produite qui est colossale et donc très délicate et fastidieuse à traiter. Cela nécessite

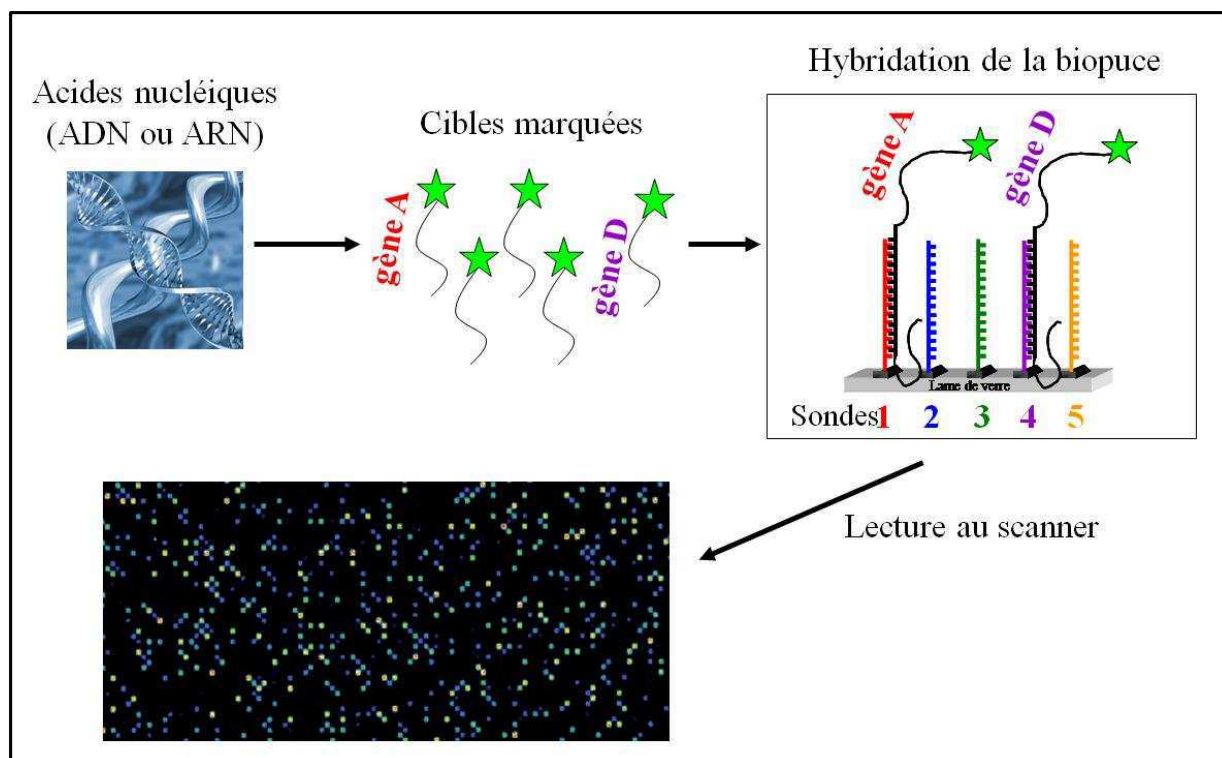


Figure 28 : Principe des biopuces ADN. Les cibles marquées par un fluorochrome s'hybrident spécifiquement avec les sondes qui leur sont complémentaires puis l'analyse de l'image obtenue permet de déterminer quels sont les gènes présents dans l'échantillon.

obligatoirement l'implication de plusieurs groupes de recherche et des moyens de calcul, de traitement et de stockage conséquents. Des approches alternatives et complémentaires peuvent donc être privilégiées, comme les biopuces ADN, qui permettent maintenant d'aborder les problématiques d'écologie microbienne pour des environnements complexes grâce à des outils également qualifiés de « haut débit » et qui nécessitent des temps de traitements de données beaucoup plus courts.

5.2. Les biopuces ADN

Les biopuces ADN sont nées de la rencontre de la microélectronique, des microsystèmes et de la biologie. Apparues au milieu des années 90, après le séquençage complet des premiers génomes, elles ont été initialement mises au point pour l'étude simultanée de l'expression de tous les gènes d'un organisme (Schena *et al.*, 1995). Cette technologie innovante est basée sur la reconnaissance entre des sondes (ADNg, produits PCR, ADNc ou oligonucléotides) fixées sur un support solide (lame de verre) et leurs cibles (produits PCR, ADNg ou ARN) marquées à l'aide d'un fluorochrome (Figure 28). Les duplex sonde/cible formés lors de l'étape d'hybridation, sont ensuite détectés à l'aide d'un scanner. Des dizaines voire des centaines de milliers de sondes différentes peuvent être fixées sur une seule biopuce, ce qui offre la possibilité d'identifier plusieurs milliers de gènes en une seule expérience. Les sondes oligonucléotidiques sont actuellement privilégiées car elles sont faciles à synthétiser et peu coûteuses (Religio *et al.*, 2002). Il existe deux principales technologies de fabrication de biopuces ADN : les biopuces *ex situ* qui nécessitent la préparation des sondes avant leur greffage sur le support solide et celles *in situ* pour lesquelles la synthèse des sondes est directement réalisée sur la lame de verre. Le choix entre ces deux technologies sera dicté par la densité désirée et le coût (Dufva, 2005; Kawasaki, 2006).

Les biopuces ADN ont été utilisées dans de nombreux domaines des sciences de la vie, y compris en écologie microbienne (Zhou, 2003; Wagner *et al.*, 2007). Il en existe différentes catégories adaptées aux nombreuses problématiques de microbiologie environnementale (Gentry *et al.*, 2006). Ainsi, la première catégorie de biopuces, appelée « Whole Genome Array » (WGA), permet de cibler l'ensemble des gènes d'un ou plusieurs microorganismes et peut être utilisée pour caractériser des souches ou des *consortia* isolés de l'environnement. Pour exemple, une biopuce composée de sondes ciblant l'ensemble des gènes présents dans les génomes de quatre souches de *Dehalococcoides* a été appliquée pour caractériser le *consortium* ANAS capable de dégrader le TCE jusqu'en éthylène (Lee *et al.*,

2011). Cependant, l'utilisation de ces biopuces pour l'étude *in situ* d'échantillons environnementaux est généralement limitée du fait de la complexité des communautés microbiennes composées majoritairement de souches pour lesquelles il n'existe pas de données sur leur génome. Il est alors plus judicieux d'utiliser des biopuces phylogénétiques (« Phylogenetic Oligonucleotide Array », POA) ou fonctionnelles (« Functional Gene Array », FGA), ciblant respectivement les biomarqueurs phylogénétiques et fonctionnels (Wagner *et al.*, 2007). La difficulté majeure pour l'élaboration de ces outils concerne l'étape de détermination et de sélection des sondes qui est dépendante de l'approche développée (WGA, POA ou FGA) et de la nature des échantillons à analyser (produit PCR, ADNg/ARN de souches ou de l'environnement). Ainsi, la pertinence de la caractérisation des communautés microbiennes repose essentiellement sur l'efficacité de ces dernières. Pour cela, elles doivent être (i) très sensibles pour permettre la détection de cibles même peu abondantes, (ii) très spécifiques pour ne reconnaître que leurs cibles sans entraîner d'hybridations aspécifiques, et (iii) le plus uniforme possible, toutes les sondes étant hybridées simultanément et dans les mêmes conditions (Loy and Bodrossy, 2006; Wagner *et al.*, 2007).

5.2.1. Les biopuces phylogénétiques (POA)

Les différents outils moléculaires développés pour identifier et classer les communautés procaryotiques ciblent principalement la séquence du biomarqueur ADNr 16S. La base de données servant à tester la spécificité des sondes doit donc contenir l'ensemble des séquences d'ARNr 16S disponibles ou caractéristiques de l'environnement étudié. Des bases de données dédiées comme SILVA (Pruesse *et al.*, 2007), Greengenes (DeSantis *et al.*, 2006) et RDP (Ribosomal Database Project) (Cole *et al.*, 2009) sont d'ailleurs maintenant accessibles. Les régions hautement conservées de l'ADNr 16S servent à élaborer des sondes ciblant les niveaux taxonomiques élevés, comme la famille, l'ordre ou la classe, et les régions hypervariables pour discriminer les microorganismes à un niveau plus fin comme le genre ou l'espèce (Huyghe *et al.*, 2008; Liles *et al.*, 2010).

La première POA appliquée à la microbiologie environnementale était composée de neuf sondes (Guschin *et al.*, 1997). Depuis, de nombreuses POA ont été proposées. Ces biopuces, généralement construites à l'aide de sondes de taille comprise entre 18 et 25 mers, trouvent de nombreuses applications pour l'étude :

- **de groupes bactériens spécifiques de différents niveaux taxonomiques** comme ceux appartenant à la division des *Acidobacteria* (Liles *et al.*, 2010) et des *Proteobacteria*

(Sanguin *et al.*, 2006b; Sanguin *et al.*, 2006a), à l'ordre des *Rhodocyclales* (Loy *et al.*, 2005) et des *Actinomycetales* (Kyselkova *et al.*, 2008), ou encore au genre *Burkholderia* (Schonmann *et al.*, 2009). Par ailleurs, la biopuce ciblant les *Proteobacteria* (Sanguin *et al.*, 2006b; Sanguin *et al.*, 2006a) a récemment été enrichie avec des sondes ciblant de nouveaux taxons et utilisée pour décrire les communautés microbiennes présentes au sein de microcosmes de sols pollués par différentes concentrations en TCE (Nemir *et al.*, 2010). Ceci a permis de montrer qu'à partir des niveaux de contaminations de l'ordre d'1 ppm, la structure de la communauté était significativement affectée.

- **de communautés microbiennes d'environnements** comme des sols pollués (Neufeld *et al.*, 2006) ou la rhizosphère du blé (Sanguin *et al.*, 2009).
- **de groupes fonctionnels particuliers** comme les sulfato-réducteurs (Loy *et al.*, 2002), les bactéries nitrifiantes (Kelly *et al.*, 2005) ou encore les *Archaea* méthanogènes (Franke-Whittle *et al.*, 2009).

En plus de ces biopuces que l'on pourrait qualifier de « dédiées », des outils plus généralistes ont également vu le jour. L'exemple de POA le plus avancé est la PhyloChip, composée de presque 500 000 sondes (Brodie *et al.*, 2006) et ciblant près de 9000 OTUs (Operational Taxonomic Units). Cet outil couvre la quasi-totalité de la diversité procaryotique répertoriée dans les bases de données. Ainsi, son utilisation permet d'obtenir une vue d'ensemble des microorganismes connus, présents au sein d'une communauté microbienne comme le prouvent les études portant sur un échantillon atmosphérique d'une zone urbaine (Brodie *et al.*, 2007), des sols de prairies (Cruz-Martinez *et al.*, 2009; DeAngelis *et al.*, 2009), d'anciennes mines contaminées (Rastogi *et al.*, 2010b; Rastogi *et al.*, 2010a) ou des sédiments de rivière contaminés par les métaux (Rastogi *et al.*, 2011). Cependant, puisque la grande majorité des microorganismes de l'environnement est toujours inconnue (Amann *et al.*, 1995; Pace, 1997), les informations obtenues par cette approche ne reflètent qu'une partie de la diversité présente dans les écosystèmes. C'est pourquoi, il paraît nécessaire de mettre en place des approches dites « exploratoires » qui pourront appréhender la diversité dans sa globalité et sans *a priori* sur les séquences ciblées. Récemment, un nouveau logiciel de sélection de sondes exploratoires pour POA, PhylArray, a été développé dans cette optique au sein de notre équipe de recherche (Militon *et al.*, 2007).

Si les POA apportent des informations pertinentes sur la composition de la microflore présente dans un environnement donné, elles sont en revanche inadaptées pour identifier leurs véritables capacités métaboliques. Ceci peut être problématique lorsque les différents membres d'un groupe bactérien présentent des capacités métaboliques différentes mais ne

peuvent être différenciés à partir des signatures de leur marqueur phylogénétique, comme c'est le cas des différentes souches appartenant au genre *Dehalococcoides* (voir paragraphe 3.1). C'est pourquoi, pour l'étude de processus métaboliques particuliers, comme ceux intervenant dans la bioremédiation des chloroéthènes, il est préférable d'avoir recours aux biopuces de type FGA, qui ciblent directement les gènes impliqués dans les processus métaboliques d'intérêt (He *et al.*, 2008).

5.2.2. Les biopuces fonctionnelles (FGA)

Les gènes ciblés par les biopuces FGA sont choisis selon des critères très précis : (i) ils doivent coder pour une enzyme clé dans la voie métabolique ciblée ; (ii) ils doivent présenter des régions suffisamment conservées pour permettre la détermination de sondes couvrant un maximum de variants de séquences mais aussi suffisamment variables pour discriminer les gènes codant pour des enzymes différentes ; (iii) ils doivent être représentés par un maximum de séquences dans les bases de données afin de disposer d'assez d'informations pour satisfaire la condition (ii). Pour la conception de FGA, les sondes oligonucléotidiques longues (50 à 70 mers) sont privilégiées pour leur meilleure sensibilité et parce que leur spécificité est suffisante au regard de la variabilité que possèdent généralement les gènes fonctionnels entre eux (Gentry *et al.*, 2006). En effet, il a été estimé que l'utilisation d'oligonucléotides de 50 mers permet de différencier des cibles possédant en général moins de 88% d'identité avec leur sonde, et de conserver une sensibilité suffisante pour détecter des cibles présentes dans 5 à 10 ng d'ADNg seul ou dans 50 à 100 ng d'ADNg en mélange (Rhee *et al.*, 2004). Ces seuils de détection correspondraient à la mise en évidence de cibles ADN provenant seulement de 10 cellules ou de 5 % des populations présentes dans une communauté bactérienne (Gentry *et al.*, 2006). De plus, il a été montré que pour des valeurs de 1 à 100 ng d'ADNg (pur ou en mélange), il existait une relation linéaire entre l'intensité du signal et la concentration de la cible ce qui signifie qu'une analyse semi quantitative serait possible pour cette gamme (Wu *et al.*, 2001).

Comme pour les POA, il existe des FGA dédiées ciblant un ou plusieurs métabolismes particuliers et d'autres plus généralistes ciblant de nombreux gènes impliqués dans la plupart des réactions biochimiques et cycles biogéochimiques. Les FGA dédiées construites ces dernières années sont particulièrement bien adaptées pour répondre à des questions biologiques précises comme celles en relation avec les cycles du méthane (Bodrossy *et al.*, 2003; Stralis-Pavese *et al.*, 2011), de l'azote (Taroncher-Oldenburg *et al.*,

2003; Ward *et al.*, 2007; Duc *et al.*, 2009) et du soufre (Rinta-Kanto *et al.*, 2011), avec la dégradation de polluants comme le benzène (Iwai *et al.*, 2007; Iwai *et al.*, 2008) ou encore avec les mécanismes de virulence bactérienne (Jaing *et al.*, 2008). Récemment, une FGA particulièrement intéressante pour l'étude des procédés de bioremédiation a été mise au point : la *BiodegPhyloChip* (Pathak *et al.*, 2011). Construite pour détecter 1057 gènes impliqués dans la dégradation de 133 polluants chimiques, elle a été utilisée pour caractériser les capacités métaboliques des microflore de plusieurs sites contaminés. Le principal inconvénient de cet outil est qu'il ne procure que des informations partielles sur les métabolismes ciblés, les sondes ne détectant ni tous les gènes impliqués dans ces processus, ni l'ensemble des variants de chaque gène. Ainsi, pour la voie de déchloration réductrice, seules des sondes ciblant les séquences des gènes *pceA* identifiés pour les genres *Dehalobacter* et *Desulfitobacterium* sont présentes.

Les différentes versions de la GeoChip, développées par le groupe du Dr. J. Zhou (pour revue : He *et al.*, 2011), sont les principales FGA généralistes actuellement disponibles. Plusieurs évolutions successives ont été proposées (Rhee *et al.*, 2004; Tiquia *et al.*, 2004; He *et al.*, 2007b; He *et al.*, 2010) permettant de passer d'un outil ciblant 2402 gènes (GeoChip 1.0) à un outil beaucoup plus élaboré ciblant maintenant 57 000 gènes (GeoChip 3.0). L'optimisation de cette biopuce semble se poursuivre puisque le groupe annonce déjà l'arrivée prochaine de la version 4.0 (He *et al.*, 2011) qui devrait permettre la détection de 152 000 gènes pour couvrir non seulement les cycles du carbone, de l'azote, du soufre et du phosphate, la réduction des métaux, la résistance aux métaux, la dégradation des contaminants organiques, mais également les voies de réponses aux stress, les phages et les éléments impliqués dans la virulence bactérienne. Ces FGA sont actuellement les plus utilisées en écologie microbienne. Elles ont notamment permis de caractériser la diversité fonctionnelle de sites pollués par de l'arsenic (Xiong *et al.*, 2010), de l'uranium (Van Nostrand *et al.*, 2009; Xu *et al.*, 2010; Van Nostrand *et al.*, 2011) et d'autres métaux (Epelde *et al.*, 2010; Xie *et al.*, 2011), des huiles (Liang *et al.*, 2009; Liang *et al.*, 2011) mais également des composés chlorés comme les PCBs (Leigh *et al.*, 2007) et l'héxachlorobenzène (Tas *et al.*, 2009). Cependant, comme les POA, la principale limite de ces outils est leur incapacité à appréhender la diversité fonctionnelle encore non décrite. Bien que l'identification de nouveaux gènes à proprement dit soit complètement exclue avec la technologie biopuce ADN, certaines stratégies sont capables, sur la base des séquences déjà répertoriées, d'aboutir à la sélection de sondes exploratoires capables d'identifier les séquences présentes dans les bases de données internationales mais également de nouveaux variants encore jamais décrits dans la

littérature et appartenant à une même famille de gènes. Ainsi, de nouveaux logiciels de détermination de sondes pour FGA intégrant ce concept ont été élaborés. Il s'agit de Metabolic Design (Terrat *et al.*, 2010) (ARTICLE ANNEXE 1) et de HiSpOD (Dugat-Bony *et al.*, 2011), ce dernier faisant l'objet d'un chapitre complet de ce manuscrit (PARTIE III ARTICLE 2).

Conclusion Générale

Depuis plus d'un siècle, une multitude de composés organochlorés anthropogènes est produite puis rejetée dans l'environnement, contaminant l'ensemble des écosystèmes (sol, air et eau). Parmi eux figurent les chloroéthènes, considérés aujourd'hui comme des polluants majeurs des eaux souterraines et des nappes phréatiques. Cependant, au sein de ces écosystèmes, plusieurs groupes microbiens possèdent l'équipement enzymatique nécessaire à leur dégradation. Celle-ci peut être réalisée aussi bien en présence qu'en absence d'oxygène. L'utilisation de ces microorganismes dans des approches dites de « bioremédiation » est une solution de traitement particulièrement pertinente pour la réhabilitation de la plupart des sites contaminés.

Malgré ce potentiel, la caractérisation et le suivi des capacités métaboliques des populations épuratrices sont des paramètres indispensables pour la mise en œuvre de stratégies de traitement efficaces. Pour cela, de nombreux outils de génomique et de post-génomique ont été conçus ces dernières décennies. Parmi eux, les biopuces ADN (POA et FGA) sont bien adaptées pour répondre aux questions d'écologie microbienne, particulièrement dans le contexte de la bioremédiation. En effet, comparées aux autres approches « haut débit » discutées précédemment, elles offrent la possibilité d'appréhender rapidement l'étude simultanée de nombreux biomarqueurs et de nombreux échantillons tout en générant des données facilement interprétables. Elles rendent donc accessible une quantité de données suffisante pour déterminer rapidement la diversité phylogénétique et/ou fonctionnelle des communautés microbiennes au sein d'environnements pollués et comprendre le fonctionnement de l'écosystème. Cependant, malgré la disponibilité d'un grand nombre d'outils, qu'ils soient dédiés ou généralistes, la plupart ne permettent pas de répondre catégoriquement à une problématique particulière. En effet, la majorité d'entre eux n'a été élaborée que pour cibler les composantes connues de la biodiversité et cela, parfois même, de manière incomplète. Ils se révèlent donc inadaptés pour l'étude d'écosystèmes complexes, hébergeant bien souvent des populations microbiennes éloignées de celles déjà identifiées. Ainsi, la mise en place d'une approche biopuce ADN pour une application en écologie microbienne nécessite encore le développement de nouveaux outils dédiés spécifiquement à la problématique posée. Cela nécessite la maîtrise d'un grand nombre d'étapes, parmi lesquelles la détermination des sondes est considérée comme cruciale. En effet, l'un des enjeux majeurs concernant l'application de cette technologie aux problématiques environnementales est la sélection de sondes qui combinent à la fois une

bonne sensibilité et une bonne spécificité, mais surtout qui présentent un caractère exploratoire ce qui permettra de détecter aussi bien des séquences de gènes connus que des séquences encore non répertoriées dans les bases de données publiques.

PARTIE II :
DETECTION DE SEQUENCES INCONNUES A
L'AIDE DES BIOPUCES ADN : STRATEGIES DE
DESIGN DE SONDAS EXPLORATOIRES

1. Contexte

Comme le montre le chapitre consacré à la synthèse bibliographique, la structure et la fonction des communautés microbiennes présentes au sein des différents écosystèmes restent encore aujourd'hui largement méconnues, principalement du fait de leur extraordinaire diversité (Whitman *et al.*, 1998; Curtis *et al.*, 2002). Cependant, grâce à l'essor spectaculaire des techniques moléculaires, les chercheurs ont littéralement changé leur manière d'appréhender le monde microbien. De nombreux outils qualifiés de « haut débit » sont apparus et avec eux, la possibilité de produire et de traiter des quantités de données gigantesques, jusqu'alors inaccessibles, et pourtant indispensables pour comprendre le fonctionnement des écosystèmes. Parmi les technologies les plus utilisées actuellement, les biopuces ADN oligonucléotidiques sont considérées comme des outils de choix pour répondre aux problématiques d'écologie microbienne (Zhou, 2003; Wagner *et al.*, 2007). En effet, grâce à elles, il est possible de caractériser dans un échantillon environnemental, la présence et/ou l'expression de plusieurs milliers de gènes et cela au cours d'une même expérience.

La pertinence d'une expérimentation biopuce ADN pour l'étude des communautés microbiennes est entièrement dépendante des sondes sélectionnées. Outre les caractéristiques obligatoires et somme toutes assez classiques qu'elles doivent présenter (sensibilité, spécificité et uniformité), leur utilisation pour des applications en écologie microbienne nécessite un caractère supplémentaire, dit « exploratoire », pour permettre la détection de l'ensemble des populations connues mais également, celles encore jamais décrites. Or, actuellement, la plupart des stratégies de détermination de sondes proposées repose uniquement sur les informations de séquences présentes dans les bases de données internationales (Lemoine *et al.*, 2009). Les sondes sélectionnées ne ciblent, par conséquent, que les microorganismes déjà connus.

2. Objectifs

Afin d'avoir une vue globale des stratégies bioinformatiques disponibles pour la détermination de sondes et applicables aux problématiques d'écologie microbienne, nous avons réalisé un état de l'art à partir de données bibliographiques. Cette partie est présentée sous forme d'une revue scientifique, publiée dans le journal « Environmental Microbiology ». Ce travail est volontairement séparé de la partie synthèse bibliographique pour permettre une

présentation détaillée des nouvelles approches de détermination de sondes exploratoires. En effet, l'augmentation des capacités de traitement des infrastructures informatiques et les récentes évolutions des formats de biopuces (plusieurs millions de sondes par lame), ont ouvert la voie à de nouvelles stratégies basées sur le design de sondes exploratoires, offrant des perspectives stimulantes pour le développement de biopuces ADN dédiées aux problématiques d'écologie microbienne.

Minireview

Detecting unknown sequences with DNA microarrays: explorative probe design strategies

Eric Dugat-Bony,^{1,2} Eric Peyretailade,^{2,3} Nicolas Parisot,^{1,2} Corinne Biderre-Petit,^{1,2} Faouzi Jaziri,^{4,5} David Hill,^{4,5} Sébastien Rimour^{2,3} and Pierre Peyret^{2,3*}

¹Clermont Université, Université Blaise Pascal, Laboratoire Microorganismes: Génome et Environnement, BP 10448, F-63000, Clermont-Ferrand.

²UMR CNRS 6023, Université Blaise Pascal, 63000 Clermont-Ferrand, France.

³Clermont Université, Université d'Auvergne, Laboratoire Microorganismes: Génome et Environnement, BP 10448, F-63000, Clermont-Ferrand.

⁴Clermont Université, Université Blaise Pascal, LIMOS, BP 10448, F-63000 Clermont-Ferrand.

⁵UMR CNRS 6158, LIMOS, F-63173 Aubière.

*For correspondence. E-mail pierre.peyret@univ-bpclermont.fr; Tel. (+33)473405139; Fax. (+33)473407670.

Running title: Explorative probes for environmental microarrays

Summary

Designing environmental DNA microarrays that can be used to survey the extreme diversity of microorganisms existing in nature, represents a stimulating challenge in the field of molecular ecology. Indeed, recent efforts in metagenomics have produced a substantial amount of sequence information from various ecosystems, and will continue to accumulate large amounts of sequence data given the qualitative and quantitative improvements in the next-generation sequencing methods. It is now possible to take advantage of this data to develop comprehensive microarrays by using explorative probe design strategies. Such strategies anticipate genetic variations and thus are able to detect known and unknown sequences in environmental samples. In this review, we provide a detailed overview of the probe design strategies currently available to construct both phylogenetic and functional DNA microarrays, with emphasis on those permitting the selection of such explorative probes. Furthermore, exploration of complex environments requires particular attention on probe sensitivity and specificity criteria. Finally, these innovative probe design approaches require exploiting newly available high-density microarray formats.

Received 31 March, 2011; accepted 4 July, 2011.

Introduction

The microbial world represents the most important and diverse group of organisms living on earth (Whitman *et al.*, 1998; Curtis *et al.*, 2002), comprising most of the diversity of the three domains of life defined by Woese *et al.* (1990): *Archaea*, *Bacteria*, and *Eucarya*. Furthermore, these organisms are widely distributed across many environmental habitats, even the most extreme. Their numerous enzymatic machineries have allowed them to adapt to almost every ecological niche and take advantage of any environmental condition (Øvreås, 2000; Guerrero and Berlanga, 2006). Despite our increasing knowledge of the role of microorganisms in ecosystem functioning, our current vision of the microbial world is still incomplete and several issues remain unsolved. This is partially explained (i) by the tremendous diversity of the genes and metabolisms of the existing species but also of ecological niches and (ii) by technological limits such as our inability to culture the majority of microorganisms (Amann *et al.*, 1995; Pace, 1997).

Because of this huge microbial biocomplexity, high-throughput molecular tools allowing simultaneous analyses of existing populations are greatly needed (Torsvik and Øvreås, 2002; Xu, 2006). Massive sequencing based on next-generation sequencing (NGS) technologies and microarrays are currently the most promising and complementary approaches to address these tasks (Claesson *et al.*, 2009; Roh *et al.*, 2010; van den Bogert *et al.*, 2011). Using NGS, two specific strategies can be applied: metagenomics which refers to the study of the collective genomes in a given environmental community and the 16S rDNA amplicon sequencing approach. In principle, these methods enable (i) access to the wide diversity of microbial communities, (ii) identification of unknown microorganisms, and (iii) the potential to link structure to functions (Simon and Daniel, 2009). Some limitations of metagenomics, however, have been demonstrated: for example, the huge difficulty of managing large amounts of sequence data, or the short sequence read length (400–500 bases maximum with 454 FLX Titanium instrument from Roche) which complicates contigs assembling, or the sequencing errors caused by NGS technologies (Roh *et al.*, 2010). Furthermore, Quince *et al.* (2008) estimated that detecting 90% of the richness in some hyperdiverse environments could require tens of thousands of times the current sequencing effort, which is inconceivable. Oligonucleotide microarray technologies have, however, been widely used for gene detection and gene expression quantification, and more recently, were adapted to profiling environmental communities in a flexible and easy-to-use manner (Zhou, 2003; Wagner *et al.*, 2007). These approaches can monitor the presence, or the expression, of

Table 1. Appropriate probe design software for microbial ecology studies.

Software	Applications in microbial ecology	Accessibility and user interface	URL	Reference
ARB	POA	Downloadable, standalone GUI (L, M)	http://www.arb-home.de/	Ludwig <i>et al.</i> (2004)
PRIMROSE	POA	Downloadable, GUI (L, W, M)	http://www.bioinformatics-toolkit.org/Primrose/index.html	Ashelford <i>et al.</i> (2002)
ORMA	POA, FGA	Matlab Script	Upon request	Severgnini <i>et al.</i> (2009)
PhyArray	POA	Web Interface	http://g2im.u-clermont1.fr/serimour/phyllarray/	Milition <i>et al.</i> (2007)
HPD	FGA	Downloadable, standalone GUI (W)	Not available	Chung <i>et al.</i> (2005)
ProDesign	FGA	Web Interface	http://www.uhnresearch.ca/labs/tillier/ProDesign/ProDesign.html	Feng and Tillier (2007)
HiSpOD	FGA, WGA	Web Interface	http://fc.isima.fr/~g2im/hispod/	Dugat-Bony <i>et al.</i> (2011)
Metabolic Design	FGA	Downloadable from a website, GUI (W)	ftp://195.221.123.90/	Terrat <i>et al.</i> (2010)
CommOligo (v 2.0)	FGA, WGA	Downloadable, standalone GUI (W)	http://ieg.ou.edu/software.htm	Li <i>et al.</i> (2005)
OligoWiz (v 2.0)	FGA, WGA	Downloadable client programme, GUI (L, W, M)	http://www.cbs.dtu.dk/services/OligoWiz/	Wernersson and Nielsen (2005)
ROSO	FGA, WGA	Web interface or standalone GUI (S, W, M) upon request	http://pbil.univ-lyon1.fr/rosor/Home.php	Reymond <i>et al.</i> (2004)
ArrayOligoSelector	FGA, WGA	Downloadable, command line (L)	http://arrayoligosel.sourceforge.net/	Bozdech <i>et al.</i> (2003)
OligoArray (v 2.1)	FGA, WGA	Downloadable, command line (L)	http://berry.engin.umich.edu/oligoarray2_1/	Rouillard <i>et al.</i> (2003)
OligoPicker	FGA, WGA	Downloadable, command line (L)	http://pga.mgh.harvard.edu/oligopicker/	Wang and Seed (2003)
PROBEmer	POA, FGA, WGA	Web Interface	Not available	Emrich <i>et al.</i> (2003)
YODA	FGA, WGA	Downloadable, standalone GUI (L, W, M)	Not available	Nordberg (2005)
ProbeSelect	WGA	Available upon request, command line (L)	Not available	Li and Stormo (2001)

FGA, functional gene array; GUI, graphical user interface; L, Linux; M, MacOS; POA, phylogenetic oligonucleotide array; S, SunOS; W, Windows; WGA, whole-genome array.

thousands of genes, combining qualitative and quantitative aspects in only one experiment (Tiquia *et al.*, 2004; Marcelino *et al.*, 2006; Dugat-Bony *et al.*, 2011). Furthermore, this technology appears well adapted to multi-sample comparison. Although several Whole-genome arrays (WGAs) have been developed in the last few years, Phylogenetic oligonucleotide arrays (POAs), targeting the 16S rRNA genes, as well as Functional gene arrays (FGAs), targeting key genes encoding enzymes involved in metabolic processes, are the two major approaches to assess diversity of microbial communities in the environment (Wagner *et al.*, 2007). Currently, the most comprehensive tools developed are the high-density PhyloChip, with nearly 500,000 oligonucleotide probes to almost 9,000 operational taxonomic units (OTU) (Brodie *et al.*, 2006), and the GeoChip 3.0 with ~28,000 probes covering approximately 57,000 gene variants from 292 functional gene families (He *et al.*, 2010). Whereas microarrays were demonstrated as being sufficiently sensitive, with detection of sequences representing genomic material from 0.05-5% of the total environmental community (Bodrossy *et al.*, 2003; Peplies *et al.*, 2004; Loy *et al.*, 2005; Gentry *et al.*, 2006; Marcelino *et al.*, 2006; Palmer *et al.*, 2006; Huyghe *et al.*, 2008), these methods require a sequence *a priori* to determine probes and hence allow surveys only of microorganisms with available sequences in public databases (Chandler and Jarrell, 2005; Wagner *et al.*, 2007).

The main problem that must be faced to construct oligonucleotide microarrays dedicated to microbial ecology is the probe design step. Indeed, environmental microarrays often require this step to be manually performed. Although numerous general probe design programmes are currently freely accessible for academics (for recent reviews see Lemoine *et al.*, (2009)), only few may be useful for microbial ecology applications and are listed in Table 1. This review aims to show how probe design strategies can avoid the limitation of sequence availability and make possible the detection of previously uncharacterized microbial populations present in nature. We emphasise various recent methods combining the use of both degenerate and non-degenerate oligonucleotide probes to target either 16S rRNA markers, or new proteic variants. In conclusion, we highlight other procedures and limitations which must be circumvented, to improve microarray development in terms of specificity and sensitivity.

General criteria for probe design

In silico probe design is one of the most critical step for microarray experiments because the selected oligonucleotide probe set will have to combine: i) sensitivity (*e.g.* probes

Table 2. Comparison of probe design software features.

Software	Probe length (nt)	Secondary structure	Low-complexity	GC content	T_m	ΔG	Degenerate probes	Cross-hybridization assessment	Database for specificity test
ARB	Fixed by the user (10–100)	No	No	Yes	Yes	No	No	Local alignment and thermodynamic calculations	ARB-Silva Database
PRIMROSE	Fixed by the user (3–100)	No	No	No	No	No	Yes	ND	RDP-II Database
ORMA	Fixed by the user	No	Yes	Yes	Yes	No	Yes	No	No
PhyArray	Fixed by the user (20–70)	No	No	Yes	Yes	No	Yes	BLAST and Kane's specifications	Custom non-redundant SSU rRNA database (95 Mo)
HPD	Fixed by the user (20–70)	Yes	No	Yes	Yes	Yes	No	BLAST and Kane's specifications	Input sequence dataset
ProDesign	Fixed by the user (20–70)	Yes	Yes	Yes	Yes	Yes	No	Spaced seed hashing and Kane's specifications	Input sequence dataset
HiSpOD	Fixed by the user	No	Yes	Yes	Yes	No	Yes	BLAST and Kane's specifications	EnvExBase (10Go) Complete CDS Database
Metabolic Design	Fixed by the user	No	No	No	No	No	Yes	BLAST and Kane's specifications	EnvExBase (10Go) Complete CDS Database
CommOligo (v 2.0)	Fixed by the user	Yes	Yes	Yes	Yes	No	No	Global alignment, thermodynamic calculations and Kane's specifications	Input sequence dataset
OligoWiz (v 2.0)	Fixed by the user	Yes	No	No	Yes	No	No	BLAST, Kane's specifications and thermodynamic calculations	Single organism genome
ROSO	Fixed by the user	Yes	Yes	Yes	Yes	Yes	No	BLAST	External fasta file (typically single organism genome)
ArrayOligoSelector	Fixed by the user	Yes	Yes	Yes	No	No	No	BLAST and thermodynamic calculations	External fasta file (typically single organism genome)
OligoArray (v 2.1)	Fixed by the user (15–75)	Yes	Yes	Yes	Yes	No	No	BLAST and thermodynamic calculations	External fasta file (typically single organism genome)
OligoPicker	Fixed by the user (20–100)	Yes	Yes	No	Yes	No	No	BLAST	Input sequence dataset or external fasta file (typically single organism genome)
PROBEmer	Fixed by the user	Yes	No	Yes	Yes	Yes	No	Suffix array approach	RDP (v 8.1), single organism genome or external fasta file
YODA	Fixed by the user	Yes	Yes	Yes	Yes	No	No	BLAST and Kane's specifications	External fasta file (typically single organism genome)
ProbeSelect	Fixed by the user	Yes	Yes	No	No	Yes	No	Suffix array approach and thermodynamic calculation	Single organism genome

ND, not determined.

should detect low abundance targets in complex mixtures), ii) specificity (*e.g.* probes should not cross-hybridize with non-target sequences) and iii) uniformity (*e.g.* probes should display similar hybridization behaviour) (Loy and Bodrossy, 2006; Wagner *et al.*, 2007). According to Lemoine *et al.* (2009), this process requires dealing with many parameters and currently available probe design programmes differ in the choice of criteria which are considered to select the best probe set (Table 2).

Sensitivity

The sensitivity generally increases with probe length, as the binding energy for longer probe–target hybrid complexes is typically higher and hybridization kinetics are irreversible (Hughes *et al.*, 2001; Religio *et al.*, 2002; Letowski *et al.*, 2004). For example, probes of 60 mers can detect targets with eight-fold higher sensitivity than those of 25 mers (Chou *et al.*, 2004). However, their threshold for differentiation is at 75-90% sequence similarity (Kane *et al.*, 2000; Taroncher-Oldenburg *et al.*, 2003; Tiquia *et al.*, 2004), which indicates a poor specificity (Li *et al.*, 2005). In contrast, short oligonucleotide probes are more specific, allowing discrimination of single nucleotide polymorphisms under optimal conditions, but at the cost of reduced sensitivity (Religio *et al.*, 2002). Furthermore, the formation of stable secondary self-structures like stem-loops, hairpins and probe-to-probe dimerization by the probes or their targets is another crucial factor which must be considered to minimize loss of microarray sensitivity. However, despite a good knowledge of the thermodynamic properties of nucleic acid duplex formation and dissociation in solution (SantaLucia *et al.*, 1996) and the availability of several algorithms like Mfold (Zuker, 2003) or Hyther (Bommarito *et al.*, 2000) for their accurate prediction, these calculations should be treated cautiously in the microarray context due to the limited knowledge on the thermodynamics of hybridization at solid-liquid interfaces (Pozhitkov *et al.*, 2006; Pozhitkov *et al.*, 2007).

Specificity

The specificity of microarray hybridization is one of the main effectors of the result quality (Kane *et al.*, 2000; Evertsz *et al.*, 2001; Koltai and Weingarten-Baror, 2008). Therefore, it is crucial that oligonucleotide probes must be unique with respect to all non-target sequences. To check probe specificity, software usually use results produced by algorithms such as BLAST or suffix array method, to search for cross-hybridization against

databases constructed in accordance with the microarray application. In this step, potential cross-hybridization prediction are usually based on Kane's recommendations (probe should not have a total percent identity >75–80% with a non-target sequence, or contiguous stretches of identity >15 nt with a non-target sequence) (Kane *et al.*, 2000) or thermodynamics calculations (duplex's stability between the probe and the nontarget-sequence). Moreover, low-complexity regions such as those containing long homopolymers may also contribute to affect probe specificity and must therefore be avoided for probe design (Wang and Seed, 2003; Leparc *et al.*, 2009).

Uniformity

Because microarray technology relies on the simultaneous hybridization of many probes under the same conditions (salt concentration, temperature, etc) it is important to ensure that the selected probes have thermodynamic behaviours as uniform as possible (Loy and Bodrossy, 2006; Wagner *et al.*, 2007). The easiest way to achieve this is to select probes with homogeneous structural properties such as probe length, G+C content, melting temperature (T_m) or binding capacities (ΔG).

Characterization of environmental species with POAs

The classical way to characterise members of complex bacterial communities relies on the small subunit ribosomal RNA gene (16S rRNA) analysis. This target is particularly well adapted to phylogenetic studies since it contains highly conserved and variable moieties permitting reliable and detailed bacterial classification. Moreover, the advent of many PCR-based approaches, as well as sequencing projects, has led to the explosion of 16S rRNA gene sequences now available in major specialized sequence repositories, such as SILVA (Pruesse *et al.*, 2007), Greengenes (DeSantis *et al.*, 2006) and the Ribosomal Database Project (RDP) (Cole *et al.*, 2009).

In order to rapidly survey prokaryotic communities present in complex environments high throughput tools have been developed, such as POAs using the SSU rRNA biomarker (Wilson *et al.*, 2002; Brodie *et al.*, 2006; Palmer *et al.*, 2006; DeSantis *et al.*, 2007). The main obstacle in designing a POA, however, is potential cross-hybridization. In many cases, the 16S rRNA genes of the type species are too conserved to allow the design of discriminatory

probes (Bae and Park, 2006). To circumvent this problem, a hierarchical design allows probing for microbial taxa at different phylogenetic levels (Huyghe *et al.*, 2008; Liles *et al.*, 2010), providing information on the presence or absence of the branches and the twigs on the Tree of Life.

Probe design for POA

Both fully-automated software and manual approaches have been developed to design POAs, taking into account the main criteria for efficient probe design, which are sensitivity and specificity. Currently, three programmes have been developed to work with structured data for retrieving and analyzing sequences from dedicated databases and to operate a phylogenetic probe design targeting the 16S rRNA gene.

The first programme is the Probe Design tool included in the ARB programme package (Ludwig *et al.*, 2004) which is commonly used to select 10–100 mer oligonucleotides. The first step in the programme consists of the target group selection. Secondly, the algorithm identifies unique sequence stretches which could serve as target sites, and subsequently returns a sorted list of potential oligonucleotides. Thirdly, the suggested probes can be matched against all sequences in the database using the Probe Match software programme. ARB also proposes different sets of predefined probes, each targeting distinct phylogenetic groups. It has been widely used to develop low-density custom made POAs, containing up to a few hundred oligonucleotide probes. These probes usually target either restricted microorganism groups known to perform a specific metabolism (Loy *et al.*, 2002; Kelly *et al.*, 2005; Franke-Whittle *et al.*, 2009), or belonging to a specific taxon (Castiglioni *et al.*, 2004; Lehner *et al.*, 2005; Loy *et al.*, 2005; Kyselkova *et al.*, 2008; Schonmann *et al.*, 2009; Liles *et al.*, 2010), or living in a habitat/ecosystem of particular interest (Neufeld *et al.*, 2006; Sanguin *et al.*, 2009). To illustrate this purpose, Sanguin *et al.* (2009) identified multiple changes in rhizobacterial community composition associated with the decline of take-all disease of wheat caused by the soil-borne fungus *Gaeumannomyces graminis* by using a taxonomic 16S rRNA-based microarray targeting both *Bacteria*, *Archaea* and the OP11 and OP2 candidate divisions. ARB has also been used to construct phylogenetic microarrays based on other biomarkers such as protein coding genes (Bodrossy *et al.*, 2003; Duc *et al.*, 2009).

The second programme is the PRIMROSE programme (Ashelford *et al.*, 2002) which uses standard or custom databases, and allows the design of degenerate probes. Initially, a

multiple alignment is produced using all the different sequences representing a given taxon. Every probe is subsequently tested against all the sequences of the initial database, to characterize potential cross-hybridizations, and to verify good coverage of the targeted taxon. Although this tool was developed to identify both phylogenetic probes and primers, it has been mainly applied to PCR-based and FISH (fluorescent in situ hybridization) approaches (Rusch and Amend, 2004; Yu *et al.*, 2005; Feldhaar *et al.*, 2007; Boeckaert *et al.*, 2008; Klitgaard *et al.*, 2008; Muhling *et al.*, 2008; Gittel *et al.*, 2009; Fraune *et al.*, 2010; Bers *et al.*, 2011). Few applications of POAs using PRIMROSE have been reported. Blaskovic and Barak (2005) reported the development of a user-friendly chip to specifically detect tick-borne bacteria responsible of human and animal diseases.

Nevertheless, neither of these two applications is built specifically for the determination of discriminating positions within a set of very similar sequences. The third programme is ORMA (Oligonucleotide Retrieving for Molecular Applications), which represents a good alternative solution (Severgnini *et al.*, 2009). This programme designs and selects oligonucleotide probes for molecular application experiments on sets of highly similar sequences. Although it was first applied to the design of probes targeting 16S rRNA genes, this software can be used on any set of highly correlated sequences, such as new potential phylogenetic biomarkers. Using this programme, Candela *et al.* (2010) designed the HTF-Microbi.Array allowing high taxonomic level fingerprinting of the human intestinal microbial community.

In parallel, other computational approaches not implemented under fully-automated software were developed to design high-density POAs (thousands of oligonucleotide probes) allowing a comprehensive screening for all known bacterial or archaeal taxa with a single microarray (Wilson *et al.*, 2002; DeSantis *et al.*, 2007). These approaches rely on sophisticated algorithms for the design of a multitude of probes and for the analyses of highly complex hybridization patterns. The best example is the PhyloChip developed by Brodie *et al.* (2006) which contains 500,000 probes based on the Affymetrix GeneChip platform. This tool is able to simultaneously identify thousands of taxa present in an environmental sample and has been applied to characterize prokaryotic communities from ecosystems such as urban atmosphere (Brodie *et al.*, 2007), grassland soils (Cruz-Martinez *et al.*, 2009; DeAngelis *et al.*, 2009), Antarctic soils (Yergeau *et al.*, 2009), mining-impacted soils (Rastogi *et al.*, 2010b; Rastogi *et al.*, 2010a), metal-contaminated river sediments (Rastogi *et al.*, 2011), terrestrial volcanic glasses (Kelly *et al.*, 2010), rhizosphere of potato (Weinert *et al.*, 2011), citrus leaf (Sagaram *et al.*, 2009), endotracheal aspirates from patients colonized by *Pseudomonas*

aeruginosa (Flanagan *et al.*, 2007), and pearly-eyed thrasher eggs (Shawkey *et al.*, 2009). Recently, due to increased interest in microbes of human and animal gastrointestinal tracts, a number of high-density microarrays were also developed to study the composition and activity of intestinal microbiota (Palmer *et al.*, 2006; Paliy *et al.*, 2009; Rajilic-Stojanovic *et al.*, 2009).

The main limitation of all the strategies proposed for 16S rRNA probe design is that they only ensure the survey of known microorganisms with available sequences in public databases. Unfortunately, the vast majority of microbial species is still unidentified and, therefore, is not represented by sequences in public ribosomal rRNA databases. A major challenge for the future is improvement of microarray technology to, in part, rely on new strategies for the design of explorative probes targeting sequences which have not yet been described.

Explorative probe design strategies for POA

The “multiple probe concept” consists of using several probes targeting an organism at similar and different phylogenetic/taxonomic levels. Designing probes using this concept significantly reduces the risk of misidentification, and often allows discrimination of bacteria down to the species level (Ludwig *et al.*, 1998; Loy and Bodrossy, 2006; Schliep and Rahmann, 2006; Huyghe *et al.*, 2008; Schonmann *et al.*, 2009; Liles *et al.*, 2010). Arrays constructed using this concept may ensure the detection of unknown taxa using probes defined from known higher phylogenetic levels. Since such probes are strictly complementary to available sequences, however, they do not harbour the explorative power to detect microorganisms with uncharacterized phylogenetic signatures.

Currently, the only software dedicated to POAs offering the possibility of designing explorative probes, is the PhylArray programme (Milton *et al.*, 2007). This algorithm generates 16S rRNA probes to globally monitor known and unknown bacterial communities in complex environments. The first step in the design is the extraction of all available sequences corresponding to a given taxon from a custom 16S rRNA curated database. Secondly, a multiple sequence alignment is performed using the ClustalW algorithm (Thompson *et al.*, 1994). Thirdly, a degenerate consensus sequence is produced taking into account sequence variability at each position, which allows the selection of degenerate probes. Finally, all combinations from each degenerate probe are checked for cross-

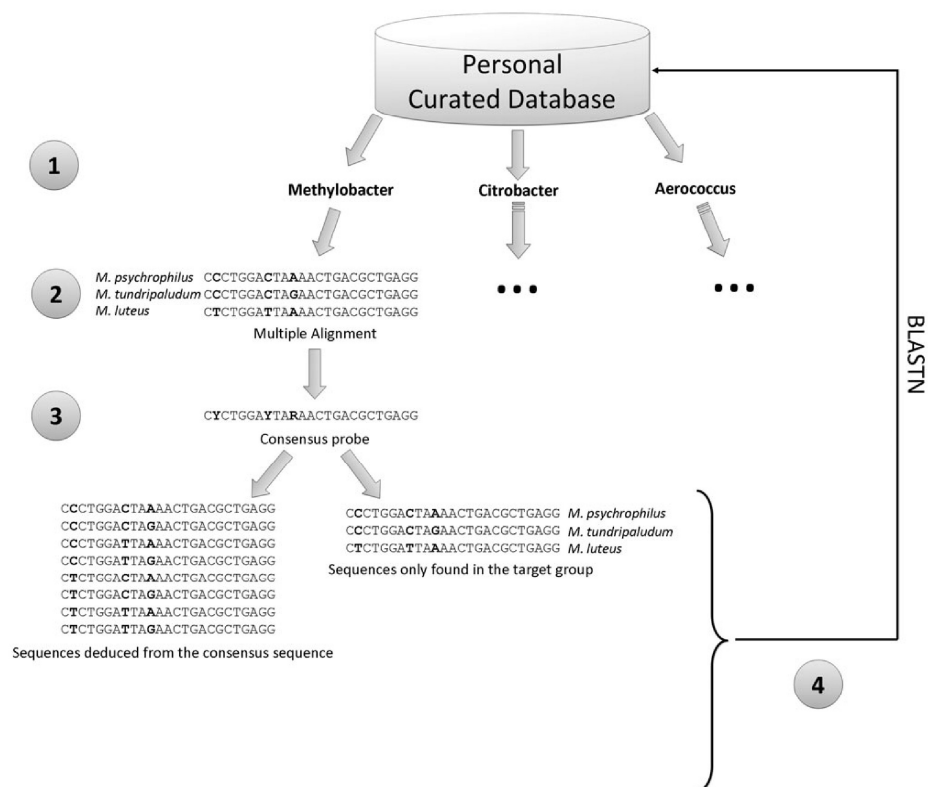


Fig. 1. PhylArray programme workflow. PhylArray programme is composed of four steps: (i) sequence extraction for each taxon, (ii) multiple sequence alignment, (iii) degenerate consensus sequence production and probe selection and (iv) specificity tests against the 16S rRNA database.

hybridization against the 16S rRNA database. Among the combinations derived from each degenerate probe, some correspond to sequences not previously included in public databases (Fig. 1). They should, therefore, allow for the exploration of the as yet undescribed fraction of environmental microbial communities. Moreover, comparative experimental evaluations indicate that probes designed with PhylArray yield a higher sensitivity and specificity than those designed with the PRIMROSE and ARB strategies (Militon *et al.*, 2007). Recently, a microarray designed with the PhylArray strategy has been employed to evaluate the bacterial diversity in two different soils (Delmont *et al.*, 2011). The authors highlighted the significant influence of several parameters like sampling depth or DNA extraction protocols on the biodiversity estimation.

Detection of functional signatures for FGA design

Assessing the metabolic potential of microorganisms in natural ecosystems is an interesting goal in microbial ecology. In fact, some authors estimate that individual environmental samples, like soil, may contain between 10^3 and 10^7 different bacterial genomes (Curtis *et al.*, 2002; Gans *et al.*, 2005), each of them harbouring thousands of genes. In this context, high-density oligonucleotide FGAs provide the best high throughput tools to access this tremendous genetic content (He *et al.*, 2008). GeoChips, composed of 50 mer probes designed with CommOligo (Li *et al.*, 2005), are currently the most comprehensive FGAs. Indeed, these microarrays have evolved over several generations and now target key genes involved in most microbial functional processes such as carbon, nitrogen, phosphorus and sulfur cycles, energy metabolism, antibiotic resistance, metal resistance and organic contaminant degradation (Rhee *et al.*, 2004; He *et al.*, 2007; He *et al.*, 2010; He *et al.*, 2011). However, being able to encompass the full diversity of gene family sequences encountered in nature, described in databanks or not, is still one of the most difficult challenges for the future. Most FGAs described to date only monitor sequences available in databases and, therefore, cannot appraise the unknown part of the microbial gene diversity present in complex environments. A more extensive coverage of the probe set is, therefore, crucial and designing explorative probes represents a pertinent and essential approach.

Characterization of new functional signatures from nucleic sequence alignment

Many probe design programmes are currently freely accessible for academics (for recent reviews see Lemoine *et al.*, (2009)). Most of them were developed for use on single-genome datasets, and hence, are limited to the determination of probes targeting specific gene sequences (Table 2). In contrast, few strategies offer the opportunity to design probes allowing a broad coverage of multiple sequence variants for a given gene family.

With the availability of more and more sequences corresponding to functional genes (complete genome sequencing and environmental studies from specific functional markers), new programmes have been developed in the last decade taking into account this wide diversity. Hierarchical Probe Design (HPD) was the first programme dedicated to functional oligonucleotide determination based on the concept of cluster-specific probes (Chung *et al.*, 2005). The first step of the programme consists of the alignment and hierarchical clustering of input sequences in order to generate all possible candidate probes. The optimal probe set is subsequently determined according to probe quality criteria, including cluster coverage, specificity, GC content and hairpin energy. Although this tool is not explorative, it automatically produces probes against all nodes of the clustering tree, providing an extensive coverage of known variants from a conserved functional gene. Using this programme, Rintakanto *et al.* (2011) developed a taxon-specific microarray targeting sulfur-related gene transcription in members of *Roseobacter* clade, using data from 13 genome sequences. This FGA consisted of 1578 probes to 431 genes and was applied to the study of diverse natural *Roseobacter* communities. The results revealed that dimethylsulfoniopropionate was not preferred over other organic carbon and sulfur substrates by these populations.

ProDesign, developed by Feng and Tillier (2007), uses similar clustering methods with the aim of detecting all members of a same gene family in environmental samples. But, unlike HPD, this software uses spaced seed hashing, rather than a suffix tree algorithm, in order to benefit from permitted mismatches between a probe and its targets, and ensures the re-clustering of groups for which no probe was found. This results in a significant improvement in sequence coverage. As with HPD, however, this tool does not provide probes targeting uncharacterized nucleic acid sequences. In addition, to the best of our knowledge, no application using this design strategy has been reported in literature.

Although both of these strategies allow a wider range of sequence variants to be covered, and, therefore, appear best suited to describe microbial communities from complex environments, their main drawbacks are their inability to generate explorative probes and the

absence of specificity tests (*i.e.* searching for potential cross-hybridizations) against large databases representative of microbial diversity. Recently, an efficient functional microarray probe design algorithm, called HiSpOD (High Specific Oligo Design), was proposed to overcome this problem (Dugat-Bony *et al.*, 2011). It is particularly useful for studying microbial communities in their environmental context. HiSpOD takes into account classical parameters for the design of effective probes (probe length, T_m, GC%, complexity) and combines supplemental properties not considered by previous programmes. Firstly, it can allow for the design of degenerate probes for gene families after multiple alignments of nucleic sequences belonging to the same gene family, and the production of consensus sequences. All combinations deduced from these degenerate probes are then divided into two groups. The first corresponds to specific probes for sequences available in databanks, and the second to explorative probes which represent potential new signatures not corresponding to any previously described microorganisms (Fig. 2A). Both the probe sets covering the most likely gene sequence variants and those covering new combinations not yet deposited in databanks are created based on multiple mutation events already identified. Secondly, the specificity of all selected probes is checked against a large formatted database dedicated to microbial communities, the EnvExBase (Environmental Expressed sequences dataBase) composed of all coding DNA sequences (CDSs) from Prokaryotes (PRO), Fungi (FUN) and Environmental (ENV) taxonomic divisions of the EMBL databank, in order to limit cross-hybridizations. To validate this strategy, a microarray focusing on the genes involved in chloroethene solvent biodegradation was developed as a model system and enabled the identification of active cooperation between *Sulfurospirillum* and *Dehalococcoides* populations in the decontamination of a polluted groundwater (Dugat-Bony *et al.*, 2011).

Use of protein sequence signatures for probe design

Unlike the strategies outlined above, a number of new strategies have been proposed to initiate probe design not from nucleic acid sequences, but from conserved peptidic regions, in order to survey all potential nucleic acid variants.

The first strategy based on this principle was described by Bontemps *et al.* (2005) and called CODEHMOP (for COnsensus DEgenerate Hybrid Motif Oligonucleotide Probe). It comes from an adaptation of the CODEHOP (for COnsensus DEgenerate Hybrid Oligonucleotide Primer) PCR primer design strategy, originally developed to identify distantly-related genes encoding proteins that belong to known families (Rose *et al.*, 1998;

Rose *et al.*, 2003; Boyce *et al.*, 2009). In the CODEHMOP strategy, conserved amino acid motifs are identified from multiple alignments of protein sequences. Then, all possible nucleic combinations (15-21 nucleotides) from the most highly conserved region (5-7 amino acids) of each protein motif are recreated and flanked by 5' and 3' fixed ends (12-15 nucleotides each), derived from the most frequent nucleotide at each position. The final probes are called "hybrids", as they consist of a variable central core, to target a larger diversity, with some nucleic combinations not corresponding to any yet described sequences, and two fixed end sequences (available in databanks) added to increase probe length. The authors used this approach to design a prototype DNA array covering all described and undescribed *nodC* (nodulation gene) sequences in bacteria, and applied it to legume nodules (Bontemps *et al.*, 2005). This strategy allowed the authors to detect new *nodC* sequences exhibiting less than 74% identity with known sequences.

The application of the CODEHMOP strategy is limited by the fact that it is not implemented into a fully-automated programme and no probe specificity test is incorporated. Nevertheless, this approach appears to be the most comprehensive way of encompassing the larger diversity of gene sequence variants potentially found for enzymes mediating a given function. Furthermore, Terrat *et al.* (2010) developed a new software programme called Metabolic Design which ensures *in silico* reconstruction of metabolic pathways, the identification of conserved motifs from protein multiple alignments, and the generation of efficient explorative probes through a simple convenient graphical interface. In this case, before the probe design stage, the user reconstructs the chosen metabolic pathway *in silico* with all substrates and products from each metabolic step. One reference enzyme for each of these steps is selected and its protein sequence extracted from a curated database (by default, Swiss-Prot) which is then used to retrieve all homologous proteins from complete databases (Swiss-Prot and TrEMBL). After selecting the most pertinent homologous sequences, they are aligned to begin the probe design stage. The amino acids are back-translated for each molecular site identified, taking into account all genetic code redundancy, to produce a degenerate nucleic consensus sequence. All degenerate probes which meet the criteria defined by the user are retained (probe length and maximal degeneracy). All the specific possible combinations for each degenerate probe are subsequently checked for potential cross-hybridizations against a representative database (*i.e.* EnvExBase as in the HiSpOD programme). Finally, an output file, listing all degenerate probes selected by the user, permits the deduction of all possible combinations and organizes them into specific probes and

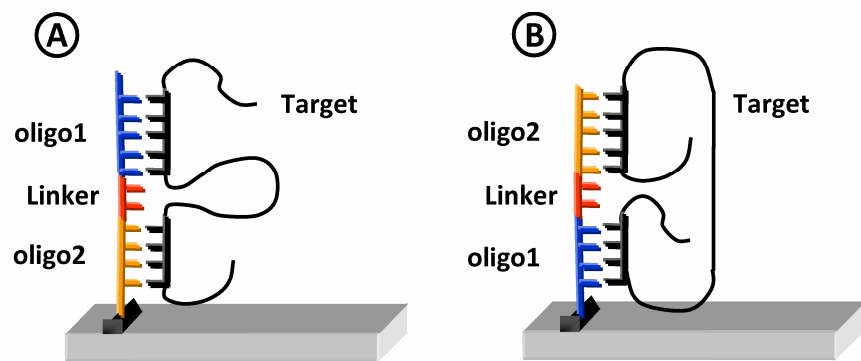


Fig. 3. Representation of the GoArrays strategy. In this strategy, two short oligonucleotide probes are concatenated with a random linker. Depending on the probes' positions, the target can form two kinds of stable loops during hybridization (A and B).

exploratory probes (Fig. 2B). The approach was validated by studying enzymes involved in the degradation of polycyclic aromatic hydrocarbons (Terrat *et al.*, 2010).

Toward circumventing microarray limitations

Despite the emergence of new design strategies, such as those presented above, the determination of a high quality probe set appears to be crucial, especially in an environmental ecology context (Liebich *et al.*, 2006; Leparc *et al.*, 2009). Although explorative potential represents a major criterion for fingerprint determination, other parameters also impact considerably on probe sensitivity and specificity, and, therefore, require particular attention (Zhou, 2003; Wagner *et al.*, 2007).

Optimization of probe size criterion

Generally, POAs employed for microbial community analysis contain short probes (typically 24-25 mers) (Brodie *et al.*, 2006; Paliy *et al.*, 2009; Rajilic-Stojanovic *et al.*, 2009), whereas FGAs are built either with short (15 to 30 mers) (Bodrossy *et al.*, 2003; Stralis-Pavese *et al.*, 2004) or long oligonucleotides (40 to 70 mers) (Kane *et al.*, 2000; Religio *et al.*, 2002; He *et al.*, 2007). The main limitation of microarrays based on short oligonucleotide probes, therefore, is the need to use, in most cases, PCR-amplified targets to ensure enrichment and thereby increase sensitivity, but this also introduces an inherent PCR bias (Suzuki and Giovannoni, 1996; Peplies *et al.*, 2004; Vora *et al.*, 2004).

An alternative approach to design oligonucleotide probes which combines excellent specificity with a potentially high sensitivity, is the use of the GoArrays strategy developed by Rimour *et al.* (2005) (software available at <http://g2im.u-clermont1.fr/serimour/goarrays.html>). In this approach, the oligonucleotide probe consists of the concatenation of two short subsequences which are complementary to disjointed regions of the target, with an insertion of a short random linker (*e.g.* 3-6 mer) (Fig. 3). This strategy has been shown to improve microarray efficiency for a wide range of applications (Rimour *et al.*, 2005; Zhou *et al.*, 2007; Pariset *et al.*, 2009; Kang *et al.*, 2010).

Specificity improvement using large databases

Because only a small portion of the natural microbial diversity has been identified, it is a major challenge to design appropriate probes specific to unique markers which do not cross-hybridize with similar unknown sequences (Chandler and Jarrell, 2005). Most of the currently available probe design software have been developed for non-environmental applications and performs specificity tests only against a reduced set of sequences, such as whole-genome data or specific sets of genes (Lemoine *et al.*, 2009). The study of microbial communities, however, requires dedicated databases that are as representative as possible of all non-target sequences potentially present in environmental samples. GenBank (Benson *et al.*, 2011), European Nucleotide Archive (ENA) (Leinonen *et al.*, 2011) and the DNA Data Bank of Japan (DDBJ) (Kaminuma *et al.*, 2011) are the most complete nucleic sequence databases publicly available to perform specificity tests. Dealing with such databases, however, is too time-consuming for probe design task, and, in this instance not really appropriate as some subsets of these databases correspond to sequences from organisms such as *Metazoa*, which are typically not considered in microbial ecology. Furthermore, for studies focusing on particular biomarkers, other sequence information need not to be considered.

For example, within POAs, each probe must be specific with respect to all small subunit (SSU) rRNA sequences which may be present in the sample during hybridization. Curated and dedicated secondary databases have been already constructed (Ribosomal Database Project (Cole *et al.*, 2009), Greengenes (DeSantis *et al.*, 2006) and SILVA (Pruesse *et al.*, 2007)), assembling all SSU rRNA sequences described on public databases. The differences between these databases come from the construction and update pipelines which lead to distinct sizes: SILVA (Release 104) contains 1,304,069 16S rRNA sequences, RDP (Release 10) 1,545,680 and Greengenes (03/22/2011) 855,446. These large databases, therefore, are well adapted to phylogenetic probe design. PhylArray software (Milton *et al.*, 2007) was developed before these databases were publicly available, and, therefore, uses its own highly-curated (full length and quality filtered) and automatically updated prokaryotic SSU rRNA database (122,337 sequences for the last release).

Because environmental FGAs target coding sequences (CDS), the database used for specificity tests must include all known CDSs which may be encountered in natural environments. To the best of our knowledge, EnvExBase (integrated in both HiSpOD and Metabolic Design programmes) is the first CDSs database dedicated to microbial ecology (Terrat *et al.*, 2010; Dugat-Bony *et al.*, 2011). For its construction, all annotated transcript

sequences and their associated 5' and 3' untranslated regions (UTR) in all classes of EMBL Prokaryotes (PRO), Fungi (FUN) and Environmental (ENV) taxonomic divisions, were extracted and curated to remove bad quality sequences. It represents a 9,129,323 sequence database.

The rapid growth of datasets, particularly environmental datasets, has led to an important increase in computational requirements coupled with a fundamental change in the way algorithms are conceived and designed (e.g. mpiBLAST (Darling *et al.*, 2003)). Consequently, parallel computing is essential, and algorithms must be deployed on large cluster infrastructures or computing grids, if specificity tests and alignments are to be performed with reasonable data processing times (Gardner *et al.*, 2006; Thorsen *et al.*, 2007).

Adaptation of the microarray format to the design strategy

Explorative design strategies targeting unknown sequences involve the use of degenerate probes (Bontemps *et al.*, 2005; Militon *et al.*, 2007; Terrat *et al.*, 2010; Dugat-Bony *et al.*, 2011). Consequently, the selected strategy will greatly influence the choice between the two major DNA microarray types (*ex-situ* or *in-situ*), the platform and the density (Dufva, 2005; Ehrenreich, 2006; Kawasaki, 2006). When using *in situ* synthesis microarrays, such as the Agilent, Affymetrix and NimbleGen platforms, all combinations resulting from a degenerate probe must be independently synthesized. This will exponentially increase the final number of probes for the array production (density). For instance, concerning the CODEHMOP (Bontemps *et al.*, 2005) and Metabolic Design strategies (Terrat *et al.*, 2010), since the genetic code often involves degeneracy at the third position of each codon, a 24 mer probe (targeting a seven amino acid conserved motif) will generate at least 128 combinations (assuming a minimal degeneracy rate of two for each codon). This value will reach at least 131,072 for a 51 mer probe containing 17 degenerate positions. Conversely, *ex-situ* platforms allow the degenerate probes (all combinations mixed together) to be spotted in the same location on the array and consequently reduce the total amount of features.

Other user choices may also affect the final number of probes per array. Replication is crucial to achieve reliable data for microarrays (Spruill *et al.*, 2002). Multiple replicates of the same probe provide some backup in case a feature cannot be evaluated due to technical artifacts, such as dye precipitations or dust particles. A statistical estimation has deduced that at least three replicates should be made (Lee *et al.*, 2000). Secondly, multiple probes per gene could be designed in order to increase confidence in the results (Loy *et al.*, 2002; Chou *et al.*,

Table 3. Characteristics of the main commercially distributed high-density microarrays.

Type of array	Technology	Probe length	Max features	Max plex
Spotted Arrays	Robot spotting Pre-made DNA	Any	~100 000	1
Affymetrix	Photolithography <i>in situ</i>	<100 mer (generally 25 mer)	~6 000 000	1
NimbleGen	Micro-mirrors <i>in situ</i>	<100 mer (generally 50–75 mer)	4 200 000	12
Agilent	Inkjet <i>in situ</i>	<100 mer (generally 25–60 mer)	1 000 000	8

2004) and to mask misleading signal variations whose causes (e.g. target secondary structure, probe folding, etc) are not yet fully understood (Pozhitkov *et al.*, 2007). Thirdly, some platforms, such as Affymetrix GeneChips, determine probe pairs where each probe (“match”) is accompanied by a negative control with a single differing base in the middle of the probe (“mismatch probe”) in order to discriminate between real signals and those due to non-specific hybridizations (Lipshutz *et al.*, 1999).

To address this problem of probe number, several commercial companies have proposed two major types of high-density microarrays whose main characteristics are described in Table 3: (i) *in situ* synthesized microarrays, distributed by Agilent (<http://www.chem.agilent.com>), NimbleGen (<http://www.nimblegen.com>) and Affymetrix (<http://www.affymetrix.com>), which can attain billions of probes and be physically divided into multi-arrays per slide (up to 12) to perform simultaneous analyses of several samples on a single experiment; and (ii) spotted microarrays (e.g. Arrayit (<http://www.arrayit.com>)) with a current printing capacity close to 100,000 features per microarray.

Concluding remarks and future directions

Assessing the extreme microbial diversity encountered in environmental samples represents an exciting challenge which could create a better understanding of microbial community functioning. Environmental DNA microarrays, with the opportunity to survey both known and unknown microorganisms through explorative probe design, are one of the most powerful approaches for achieving this goal. Future perspectives in this domain will be to systematically integrate this innovative concept into probe design workflows, especially by offering the possibility to design degenerate probes targeting sequence clusters. Furthermore, to efficiently recognize signals due to unknown targets, it will be particularly useful to develop automatic procedures to analyse microarray data. In addition, using explorative probe design in sequence capture approaches that couple with NGS, such as those originally developed for direct selection of human genomic loci (Albert *et al.*, 2007), could also improve this gene characterization. Indeed, sequence capture elution products should allow the full identification and characterization of new taxa when using POAs or new protein coding genes with FGAs.

The constant increase in available sequences (Cochrane *et al.*, 2009) means that databases for specificity tests must be regularly updated. As a result, probe datasets must be

re-computed as frequently as possible in order to take into account all deposited data. Nevertheless, assessing probe specificity against large databases is a time-consuming task. To overcome this problem two complementary strategies could be employed:

i) Creation of databases specific to each ecological compartment. Usually, specificity tests are not performed against a suitable subset of sequences mainly due to lack of databases for microbial ecology. Depending on the environment studied it would be more relevant to perform these tests against reduced databanks dedicated to specific ecosystems (soil, marine, freshwater, gut, etc).

ii) Parallelization of probe design algorithms. Perspectives to limit computation time are based on exploiting the computational resources available using specialized frameworks such as Message Passing Interface (MPI) or heterogeneous systems including General-purpose Processing on Graphics Processing Units (GPGPU). With the recent development of extremely fast broadband networks, it has become possible to distribute the calculations at larger and larger scales over different geographical locations (Schadt *et al.*, 2010). Cluster, grid or emerging cloud computing are all examples of shared computing resources where probe design algorithms can be deployed. Being able to improve the bioinformatics tools applied to environmental microbiology through algorithm deployment on such shared computational resources, and combining them with automatic update pipelines, are two important challenges and strategies for the future of the field of molecular ecology.

Acknowledgments

This work was supported by the grant ID 2598 from the “Agence De l’Environnement et de la Maîtrise de l’Energie” (ADEME, France); the grant ANR-07-ECOT-005-05 for the programme PRECODD Evasol from ‘Agence Nationale de la Recherche’ (ANR, France); the grant ANR-08-BIOENERGIES-0 for the programme BIOENERGIES AnaBio-H2 from ‘Agence Nationale de la Recherche’ (ANR, France); and the INSU-EC2CO programme from ‘Centre National de la Recherche Scientifique’ (CNRS, France). We thank David Tottey for reviewing the English version of the manuscript.

References

- Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X. et al. (2007) Direct selection of human genomic loci by microarray hybridization. *Nat Methods* **4**: 903-905.
- Amann, R., Ludwig, W., and Schleifer, K. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.
- Ashelford, K.E., Weightman, A.J., and Fry, J.C. (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res* **30**: 3481-3489.
- Bae, J.W., and Park, Y.H. (2006) Homogeneous versus heterogeneous probes for microbial ecological microarrays. *Trends Biotechnol* **24**: 318-323.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res* **39**: D32-37.
- Bers, K., Sniegowski, K., Albers, P., Breugelmans, P., Hendrickx, L., De Mot, R., and Springael, D. (2011) A molecular toolbox to estimate the number and diversity of *Variovorax* in the environment: application in soils treated with the phenylurea herbicide linuron. *FEMS Microbiol Ecol* **76**: 14-25.
- Blaskovic, D., and Barak, I. (2005) Oligo-chip based detection of tick-borne bacteria. *FEMS Microbiol Lett* **243**: 473-478.
- Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5**: 566-582.
- Boeckaert, C., Vlaeminck, B., Fievez, V., Maignien, L., Dijkstra, J., and Boon, N. (2008) Accumulation of trans C18:1 fatty acids in the rumen after dietary algal supplementation is associated with changes in the *Butyrivibrio* community. *Appl Environ Microbiol* **74**: 6923-6930.
- Bommarito, S., Peyret, N., and SantaLucia, J., Jr. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res* **28**: 1929-1934.
- Bontemps, C., Golfier, G., Gris-Liebe, C., Carrere, S., Talini, L., and Boivin-Masson, C. (2005) Microarray-based detection and typing of the *Rhizobium* nodulation gene *nodC*: potential of DNA arrays to diagnose biological functions of interest. *Appl Environ Microbiol* **71**: 8042-8048.
- Boyce, R., Chilana, P., and Rose, T.M. (2009) iCODEHOP: a new interactive program for designing CONsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. *Nucleic Acids Res* **37**: W222-228.
- Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., and DeRisi, J.L. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* **4**: R9.
- Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M., and Andersen, G.L. (2007) Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* **104**: 299-304.
- Brodie, E.L., Desantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L. et al. (2006) Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72**: 6288-6298.
- Candela, M., Consolandi, C., Severgnini, M., Biagi, E., Castiglioni, B., Vitali, B. et al. (2010) High taxonomic level fingerprint of the human intestinal microbiota by ligase detection reaction--universal array approach. *BMC Microbiol* **10**: 116.

- Castiglioni, B., Rizzi, E., Frosini, A., Sivonen, K., Rajaniemi, P., Rantala, A. et al. (2004) Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria. *Appl Environ Microbiol* **70**: 7161-7172.
- Chandler, D.P., and Jarrell, A.E. (2005) Taking arrays from the lab to the field: trying to make sense of the unknown. *Biotechniques* **38**: 591-600.
- Chou, C.C., Chen, C.H., Lee, T.T., and Peck, K. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res* **32**: e99.
- Chung, W.-H., Rhee, S.-K., Wan, X.-F., Bae, J.-W., Quan, Z.-X., and Park, Y.-H. (2005) Design of long oligonucleotide probes for functional gene detection in a microbial community. *Bioinformatics* **21**: 4092-4100.
- Claesson, M.J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J.R., Smidt, H. et al. (2009) Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS One* **4**: e6669.
- Cochrane, G., Akhtar, R., Bonfield, J., Bower, L., Demiralp, F., Faruque, N. et al. (2009) Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Res* **37**: D19-25.
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J. et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141-145.
- Cruz-Martinez, K., Suttle, K.B., Brodie, E.L., Power, M.E., Andersen, G.L., and Banfield, J.F. (2009) Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *ISME J* **3**: 738-744.
- Curtis, T.P., Sloan, W.T., and Scannell, J.W. (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99**: 10494-10499.
- Darling, A., Carey, L., and Feng, W. (2003) The Design, Implementation, and Evaluation of mpiBLAST. In: *4th International Conference on Linux Clusters: The HPC Revolution 2003*. San Jose, California.
- DeAngelis, K.M., Brodie, E.L., DeSantis, T.Z., Andersen, G.L., Lindow, S.E., and Firestone, M.K. (2009) Selective progressive response of soil microbial community to wild oat roots. *ISME J* **3**: 168-178.
- Delmont, T.O., Robe, P., Cecillon, S., Clark, I.M., Constancias, F., Simonet, P. et al. (2011) Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* **77**: 1315-1324.
- DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M., and Andersen, G.L. (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* **53**: 371-383.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. et al. (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* **72**: 5069-5072.
- Duc, L., Neuenschwander, S., Rehrauer, H., Wagner, U., Sobek, J., Schlapbach, R., and Zeyer, J. (2009) Development and experimental validation of a nifH oligonucleotide microarray to study diazotrophic communities in a glacier forefield. *Environ Microbiol* **11**: 2179-2189.
- Dufva, M. (2005) Fabrication of high quality microarrays. *Biomol Eng* **22**: 173-184.
- Dugat-Bony, E., Missaoui, M., Peyretailade, E., Biderre-Petit, C., Bouzid, O., Gouinaud, C. et al. (2011) HiSpOD: probe design for functional DNA microarrays. *Bioinformatics* **27**: 641-648.
- Ehrenreich, A. (2006) DNA microarray technology for the microbiologist: an overview. *Appl Microbiol Biotechnol* **73**: 255-273.

- Emrich, S.J., Lowe, M., and Delcher, A.L. (2003) PROBEmer: A web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Res* **31**: 3746-3750.
- Evertsz, E.M., Au-Young, J., Ruvolo, M.V., Lim, A.C., and Reynolds, M.A. (2001) Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques* **31**: 1182, 1184, 1186 passim.
- Feldhaar, H., Straka, J., Krischke, M., Berthold, K., Stoll, S., Mueller, M.J., and Gross, R. (2007) Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol* **5**: 48.
- Feng, S., and Tillier, E.R.M. (2007) A fast and flexible approach to oligonucleotide probe design for genomes and gene families. *Bioinformatics* **23**: 1195-1202.
- Flanagan, J.L., Brodie, E.L., Weng, L., Lynch, S.V., Garcia, O., Brown, R. et al. (2007) Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with *Pseudomonas aeruginosa*. *J Clin Microbiol* **45**: 1954-1962.
- Franke-Whittle, I.H., Goberna, M., Pfister, V., and Insam, H. (2009) Design and development of the ANAEROCHIP microarray for investigation of methanogenic communities. *J Microbiol Methods* **79**: 279-288.
- Fraune, S., Augustin, R., Anton-Erxleben, F., Wittlieb, J., Gelhaus, C., Klimovich, V.B. et al. (2010) In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proc Natl Acad Sci USA* **107**: 18067-18072.
- Gans, J., Wolinsky, M., and Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**: 1387-1390.
- Gardner, M.K., Feng, W.-c., Archuleta, J., Lin, H., and Mal, X. (2006) Parallel genomic sequence-searching on an ad-hoc grid: experiences, lessons learned, and implications. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*. SC Conference (ed.). Tampa, Florida: ACM, pp.22.
- Gentry, T., Wickham, G., Schadt, C., He, Z., and Zhou, J. (2006) Microarray Applications in Microbial Ecology Research. *Microb Ecol* **52**: 159-175.
- Gittel, A., Sorensen, K.B., Skovhus, T.L., Ingvorsen, K., and Schramm, A. (2009) Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment. *Appl Environ Microbiol* **75**: 7086-7096.
- Guerrero, R., and Berlanga, M. (2006) Life's unity and flexibility: the ecological link. *Int Microbiol* **9**: 225-235.
- He, Z., Van Nostrand, J.D., Deng, Y., and Zhou, J.Z. (2011) Development and applications of functional gene microarrays in the analysis of the functional diversity, composition, and structure of microbial communities. *Front Environ Sci Engin China* **5**: 1-20.
- He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C. et al. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67-77.
- He, Z., Deng, Y., Van Nostrand, J.D., Tu, Q., Xu, M., Hemme, C.L. et al. (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4**: 1167-1179.
- He, Z.L., Van Nostrand, J.D., Wu, L.Y., and Zhou, J.Z. (2008) Development and application of functional gene arrays for microbial community analysis. *Trans Nonferrous Met Soc China* **18**: 1319-1327.
- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W. et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**: 342-347.
- Huyghe, A., Francois, P., Charbonnier, Y., Tangomo-Bento, M., Bonetti, E.-J., Paster, B.J. et al. (2008) Novel Microarray Design Strategy To Study Complex Bacterial Communities. *Appl Environ Microbiol* **74**: 1876-1885.

- Kaminuma, E., Kosuge, T., Kodama, Y., Aono, H., Mashima, J., Gojobori, T. et al. (2011) DDBJ progress report. *Nucleic Acids Res* **39**: D22-27.
- Kane, M.D., Jatke, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**: 4552-4557.
- Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and McSweeney, C.S. (2010) Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* **16**: 2034-2042.
- Kawasaki, E.S. (2006) The end of the microarray Tower of Babel: will universal standards lead the way? *J Biomol Tech* **17**: 200-206.
- Kelly, J.J., Siripong, S., McCormack, J., Janus, L.R., Urakawa, H., El Fantroussi, S. et al. (2005) DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. *Water Res* **39**: 3229-3238.
- Kelly, L.C., Cockell, C.S., Piceno, Y.M., Andersen, G.L., Thorsteinsson, T., and Marteinsson, V. (2010) Bacterial diversity of weathered terrestrial Icelandic volcanic glasses. *Microb Ecol* **60**: 740-752.
- Klitgaard, K., Boye, M., Capion, N., and Jensen, T.K. (2008) Evidence of multiple *Treponema* phylotypes involved in bovine digital dermatitis as shown by 16S rRNA gene analysis and fluorescence in situ hybridization. *J Clin Microbiol* **46**: 3012-3020.
- Koltai, H., and Weingarten-Baror, C. (2008) Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction. *Nucleic Acids Res* **36**: 2395-2405.
- Kyselkova, M., Kopecky, J., Felfoldi, T., Cermak, L., Omelka, M., Grundmann, G.L. et al. (2008) Development of a 16S rRNA gene-based prototype microarray for the detection of selected actinomycetes genera. *Antonie Van Leeuwenhoek* **94**: 439-453.
- Lee, M.L., Kuo, F.C., Whitmore, G.A., and Sklar, J. (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* **97**: 9834-9839.
- Lehner, A., Loy, A., Behr, T., Gaenge, H., Ludwig, W., Wagner, M., and Schleifer, K.H. (2005) Oligonucleotide microarray for identification of *Enterococcus* species. *FEMS Microbiol Lett* **246**: 133-142.
- Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdano-Tarraga, A., Cheng, Y. et al. (2011) The European Nucleotide Archive. *Nucleic Acids Res* **39**: D28-31.
- Lemoine, S., Combes, F., and Le Crom, S. (2009) An evaluation of custom microarray applications: the oligonucleotide design challenge. *Nucleic Acids Res* **37**: 1726-1739.
- Leparc, G.G., Tuchler, T., Striedner, G., Bayer, K., Sykacek, P., Hofacker, I.L., and Kreil, D.P. (2009) Model-based probe set optimization for high-performance microarrays. *Nucleic Acids Res* **37**: e18.
- Letowski, J., Brousseau, R., and Masson, L. (2004) Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J Microbiol Methods* **57**: 269-278.
- Li, F., and Stormo, G.D. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* **17**: 1067-1076.
- Li, X., He, Z., and Zhou, J. (2005) Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res* **33**: 6114-6123.
- Liebich, J., Schadt, C.W., Chong, S.C., He, Z., Rhee, S.K., and Zhou, J. (2006) Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. *Appl Environ Microbiol* **72**: 1688-1691.

- Liles, M.R., Turkmen, O., Manske, B.F., Zhang, M., Rouillard, J.-M., George, I. et al. (2010) A phylogenetic microarray targeting 16S rRNA genes from the bacterial division Acidobacteria reveals a lineage-specific distribution in a soil clay fraction. *Soil Biol Biochem* **42**: 739-747.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat Genet* **21**: 20-24.
- Loy, A., and Bodrossy, L. (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clin Chim Acta* **363**: 106-119.
- Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J. et al. (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68**: 5064-5081.
- Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoecker, K., Baranyi, C. et al. (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". *Appl Environ Microbiol* **71**: 1373-1386.
- Ludwig, W., Amann, R., Martinez-Romero, E., Schönhuber, W., Bauer, S., Neef, A., and Schleifer, K.-H. (1998) rRNA based identification and detection systems for rhizobia and other bacteria. *Plant Soil* **204**: 1-19.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Kumar, Y. et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
- Marcelino, L.A., Backman, V., Donaldson, A., Steadman, C., Thompson, J.R., Preheim, S.P. et al. (2006) Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. *Proc Natl Acad Sci USA* **103**: 13629-13634.
- Milton, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D. et al. (2007) PhylArray: phylogenetic probe design algorithm for microarray. *Bioinformatics* **23**: 2550-2557.
- Muhling, M., Woolven-Allen, J., Murrell, J.C., and Joint, I. (2008) Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J* **2**: 379-392.
- Neufeld, J.D., Mohn, W.W., and de Lorenzo, V. (2006) Composition of microbial communities in hexachlorocyclohexane (HCH) contaminated soils from Spain revealed with a habitat-specific microarray. *Environ Microbiol* **8**: 126-140.
- Nordberg, E.K. (2005) YODA: selecting signature oligonucleotides. *Bioinformatics* **21**: 1365-1370.
- Øvreås, L. (2000) Population and community level approaches for analysing microbial diversity in natural environments. *Ecol Lett* **3**: 236-251.
- Pace, N.R. (1997) A Molecular View of Microbial Diversity and the Biosphere. *Science* **276**: 734-740.
- Paliy, O., Kenche, H., Abernathy, F., and Michail, S. (2009) High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray. *Appl Environ Microbiol* **75**: 3572-3579.
- Palmer, C., Bik, E.M., Eisen, M.B., Eckburg, P.B., Sana, T.R., Wolber, P.K. et al. (2006) Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res* **34**: e5.
- Pariset, L., Chillemi, G., Bongiorno, S., Romano Spica, V., and Valentini, A. (2009) Microarrays and high-throughput transcriptomic analysis in species with incomplete availability of genomic sequences. *Nat Biotechnol* **25**: 272-279.
- Peplies, J., Lau, S.C.K., Pernthaler, J., Amann, R., and Glöckner, F.O. (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton. *Environ Microbiol* **6**: 638-645.

- Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P. et al. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* **34**: e66.
- Pozhitkov, A.E., Tautz, D., and Noble, P.A. (2007) Oligonucleotide microarrays: widely applied--poorly understood. *Brief Funct Genomic Proteomic* **6**: 141-148.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188-7196.
- Quince, C., Curtis, T.P., and Sloan, W.T. (2008) The rational exploration of microbial diversity. *ISME J* **2**: 997-1006.
- Rajilic-Stojanovic, M., Heilig, H.G., Molenaar, D., Kajander, K., Surakka, A., Smidt, H., and de Vos, W.M. (2009) Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ Microbiol* **11**: 1736-1751.
- Rastogi, G., Osman, S., Vaishampayan, P.A., Andersen, G.L., Stetler, L.D., and Sani, R.K. (2010a) Microbial diversity in uranium mining-impacted soils as revealed by high-density 16S microarray and clone library. *Microb Ecol* **59**: 94-108.
- Rastogi, G., Osman, S., Kukkadapu, R., Engelhard, M., Vaishampayan, P.A., Andersen, G.L., and Sani, R.K. (2010b) Microbial and mineralogical characterizations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microb Ecol* **60**: 539-550.
- Rastogi, G., Barua, S., Sani, R.K., and Peyton, B.M. (2011) Investigation of Microbial Populations in the Extremely Metal-Contaminated Coeur d'Alene River Sediments. *Microb Ecol* **62**: 1-13.
- Religio, A., Schwager, C., Richter, A., Ansorge, W., and Valcarcel, J. (2002) Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res* **30**: e51.
- Reymond, N., Charles, H., Duret, L., Calevro, F., Beslon, G., and Fayard, J.M. (2004) ROSO: optimizing oligonucleotide probes for microarrays. *Bioinformatics* **20**: 271-273.
- Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X., and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303-4317.
- Rimour, S., Hill, D., Milton, C., and Peyret, P. (2005) GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics* **21**: 1094-1103.
- Rinta-Kanto, J.M., Burgmann, H., Gifford, S.M., Sun, S., Sharma, S., del Valle, D.A. et al. (2011) Analysis of sulfur-related transcription by Roseobacter communities using a taxon-specific functional gene microarray. *Environ Microbiol* **13**: 453-467.
- Roh, S.W., Abell, G.C.J., Kim, K.-H., Nam, Y.-D., and Bae, J.-W. (2010) Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends Biotechnol* **28**: 291-299.
- Rose, T.M., Henikoff, J.G., and Henikoff, S. (2003) CODEHOP (CONsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* **31**: 3763-3766.
- Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**: 1628-1635.
- Rouillard, J.M., Zuker, M., and Gulari, E. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* **31**: 3057-3062.

- Rusch, A., and Amend, J.P. (2004) Order-specific 16S rRNA-targeted oligonucleotide probes for (hyper)thermophilic archaea and bacteria. *Extremophiles* **8**: 357-366.
- Sagaram, U.S., DeAngelis, K.M., Trivedi, P., Andersen, G.L., Lu, S.E., and Wang, N. (2009) Bacterial diversity analysis of Huanglongbing pathogen-infected citrus, using PhyloChip arrays and 16S rRNA gene clone library sequencing. *Appl Environ Microbiol* **75**: 1566-1574.
- Sanguin, H., Sarniguet, A., Gazengel, K., Moenne-Loccoz, Y., and Grundmann, G.L. (2009) Rhizosphere bacterial communities associated with disease suppressiveness stages of take-all decline in wheat monoculture. *New Phytol* **184**: 694-707.
- SantaLucia, J., Jr., Allawi, H.T., and Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**: 3555-3562.
- Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., and Nolan, G.P. (2010) Computational solutions to large-scale data management and analysis. *Nat Rev Genet* **11**: 647-657.
- Schliep, A., and Rahmann, S. (2006) Decoding non-unique oligonucleotide hybridization experiments of targets related by a phylogenetic tree. *Bioinformatics* **22**: e424-430.
- Schonmann, S., Loy, A., Wimmersberger, C., Sobek, J., Aquino, C., Vandamme, P. et al. (2009) 16S rRNA gene-based phylogenetic microarray for simultaneous identification of members of the genus Burkholderia. *Environ Microbiol* **11**: 779-800.
- Severgnini, M., Cremonesi, P., Consolandi, C., Caredda, G., De Bellis, G., and Castiglioni, B. (2009) ORMA: a tool for identification of species-specific variations in 16S rRNA gene and oligonucleotides design. *Nucleic Acids Res* **37**: e109.
- Shawkey, M.D., Firestone, M.K., Brodie, E.L., and Beissinger, S.R. (2009) Avian incubation inhibits growth and diversification of bacterial assemblages on eggs. *PLoS One* **4**: e4522.
- Simon, C., and Daniel, R. (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* **85**: 265-276.
- Spruill, S.E., Lu, J., Hardy, S., and Weir, B. (2002) Assessing sources of variability in microarray gene expression data. *Biotechniques* **33**: 916-920, 922-913.
- Stralis-Pavese, N., Sessitsch, A., Weilharter, A., Reichenauer, T., Riesing, J., Csontos, J. et al. (2004) Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environ Microbiol* **6**: 347-363.
- Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625-630.
- Taroncher-Oldenburg, G., Griner, E.M., Francis, C.A., and Ward, B.B. (2003) Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* **69**: 1159-1171.
- Terrat, S., Peyretilade, E., Goncalves, O., Dugat-Bony, E., Gravelat, F., Mone, A. et al. (2010) Detecting variants with Metabolic Design, a new software tool to design probes for explorative functional DNA microarray development. *BMC Bioinformatics* **11**: 478.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.
- Thorsen, O., Smith, B., Sosa, C.P., Jiang, K., Lin, H., Peters, A., and Feng, W.-c. (2007) Parallel genomic sequence-search on a massively parallel system. In *Proceedings of the 4th international conference on Computing frontiers*. ACM (ed.). Ischia, Italy: pp.59-68.

- Tiquia, S.M., Wu, L., Chong, S.C., Passovets, S., Xu, D., Xu, Y., and Zhou, J. (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **36**: 664-670.
- Torsvik, V., and Øvreås, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240-245.
- van den Bogert, B., de Vos, W.M., Zoetendal, E.G., and Kleerebezem, M. (2011) Microarray Analysis and Barcoded Pyrosequencing Provide Consistent Microbial Profiles Depending on the Source of Human Intestinal Samples. *Appl Environ Microbiol* **77**: 2071-2080.
- Vora, G.J., Meador, C.E., Stenger, D.A., and Andreadis, J.D. (2004) Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl Environ Microbiol* **70**: 3047-3054.
- Wagner, M., Smidt, H., Loy, A., and Zhou, J. (2007) Unravelling Microbial Communities with DNA-Microarrays: Challenges and Future Directions. *Microb Ecol* **53**: 498-506-506.
- Wang, X., and Seed, B. (2003) Selection of oligonucleotide probes for protein coding sequences. *Bioinformatics* **19**: 796-802.
- Weinert, N., Piceno, Y., Ding, G.C., Meincke, R., Heuer, H., Berg, G. et al. (2011) PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol Ecol* **75**: 497-506.
- Wernersson, R., and Nielsen, H.B. (2005) OligoWiz 2.0--integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res* **33**: W611-615.
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* **95**: 6578-6583.
- Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A., and Andersen, G.L. (2002) High-Density Microarray of Small-Subunit Ribosomal DNA Probes. *Appl Environ Microbiol* **68**: 2535-2541.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87**: 4576-4579.
- Xu, J. (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* **15**: 1713-1731.
- Yergeau, E., Schoondermark-Stolk, S.A., Brodie, E.L., Dejean, S., DeSantis, T.Z., Goncalves, O. et al. (2009) Environmental microarray analyses of Antarctic soil microbial communities. *ISME J* **3**: 340-351.
- Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005) Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng* **89**: 670-679.
- Zhou, J. (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288-294.
- Zhou, Z., Dou, Z.-x., Zhang, C., Yu, H.-q., Liu, Y.-j., Zhang, C.-z., and Cao, Y.-j. (2007) A strategy to optimize the oligo-probes for microarray-based detection of viruses. *Virol Sin* **22**: 326-335-335.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.

Tableau 13: Comparaison des approches de méta«omiques» et de biopuces ADN pour des applications en écologie microbienne. + : approche mieux adaptée. - : approche moins bien adaptée.

Approches	Méta«omiques»	Biopuce ADN
Profondeur de l'analyse	+	-
Analyse globale d'un écosystème	+	-
Analyse ciblée d'un processus	-	+
Débit (nombre d'échantillons par expérience)	-	+
Coût par échantillon	-	+
Analyse en routine	-	+
Facilité d'interprétation des résultats	-	+
Caractérisation de nouvelles fonctions / microorganismes	+	-
Identification de nouveaux variants génétiques	+	+

3. Discussion

Pouvoir explorer l'immense biodiversité présente dans l'environnement est encore aujourd'hui l'un des enjeux majeurs de l'écologie microbienne. Les biopuces ADN possèdent désormais les atouts nécessaires pour atteindre cet objectif notamment grâce (i) aux nouveaux concepts de design de sondes dites « exploratoires » qui permettent d'accéder à la fraction inconnue des communautés microbiennes, (ii) à la puissance des outils informatiques permettant de s'assurer de la pertinence des sondes sélectionnées et (iii) au développement de nouveaux formats de biopuces ADN pouvant contenir plusieurs millions de sondes.

Devant l'extraordinaire diversité présente sur terre, les scientifiques tentent de repousser les limites de la technologie afin d'accéder à cet immense réservoir d'informations. Plusieurs stratégies, avec chacune leurs avantages et leurs inconvénients, ont ainsi vu le jour (Tableau 13). Celles basées sur les approches de métagénomique et de métatranscriptomique, qualifiées de « haut débit », permettent d'explorer les écosystèmes d'intérêt et d'identifier de nouveaux gènes ou de nouveaux variants génétiques. Elles génèrent pour cela, des quantités de données colossales qu'il est encore souvent difficile d'analyser et d'exploiter à l'échelle d'un laboratoire. D'autres stratégies, en revanche, sont basées sur le principe inverse : la diversité n'est plus recherchée directement dans l'écosystème d'intérêt, elle est, en effet, imaginée au travers des signatures nucléiques dégénérées construites en prenant en compte l'ensemble de la variabilité des séquences nucléiques ou protéiques déjà disponibles dans les bases de données. Ainsi, en considérant de nouvelles règles d'associations déduites des signatures identifiées, de nouvelles combinaisons de séquences sont créées, séquences valides au niveau génétique et donc potentiellement portées par des individus présents dans les écosystèmes, bien que jamais observées auparavant. Ce concept, que l'on peut qualifier d'« exploratoire », a d'abord été appliqué pour la détermination d'amorces PCR et validé par la caractérisation de nouvelles séquences parfois très éloignées de celles déjà connues (Rose *et al.*, 1998; Rose *et al.*, 2003). Son utilisation s'étend maintenant à d'autres outils de biologie moléculaire comme les biopuces ADN (Bontemps *et al.*, 2005). Cependant, elle nécessite le développement d'outils informatiques adaptés.

L'élaboration de logiciels pour la détermination de sondes dégénérées nécessite de prendre en compte plusieurs paramètres. En effet, chaque sonde dégénérée correspond en réalité à un ensemble de combinaisons dont le nombre peut être plus ou moins grand selon le niveau de dégénérescence (Figure 2 de l'article). Afin d'éviter toute hybridation aspécifique, la spécificité de chaque combinaison est vérifiée contre une base de données qui doit tenir

compte de toute la biodiversité potentiellement présente dans l'écosystème d'intérêt. Ainsi pour l'élaboration d'une POA, le test de spécificité de chaque sonde ADNr 16S doit être réalisé contre une base de données contenant toutes les séquences ARNr 16S existantes. De même, pour une FGA, la base de données devra contenir toutes les séquences codant pour des gènes identifiés chez des microorganismes isolés ou à partir d'échantillons environnementaux. Devant la croissance quasi exponentielle du nombre de séquences dans les bases de données généralistes (Cochrane *et al.*, 2009), ces bases spécialisées constituées autour de thématiques biologiques spécifiques deviennent également de plus en plus conséquentes. Ainsi, la recherche de similarité effectuée contre ces bases nécessite une quantité de calculs importante obligeant l'adaptation de la structure des algorithmes pour pouvoir utiliser des ressources informatiques puissantes [multiprocesseurs (SMP) et utilisation des processeurs des cartes graphiques (GPGPU)] ou pour partager les calculs à réaliser sur des architectures de type clusters ou grilles de calculs (Schadt *et al.*, 2010).

Bien que ce chapitre souligne la difficulté de mettre en place une approche biopuce ADN, surtout parce qu'il est nécessaire d'avoir recours à des logiciels efficaces pour la conception de sondes, il met également en avant la puissance de cet outil.

**PARTIE III :
DEVELOPPEMENT D'UN LOGICIEL DE
SELECTION DE SONDES
OLIGONUCLEOTIDIQUES POUR BIOPUCES
ADN FONCTIONNELLES**

1. Contexte

L'étude des capacités métaboliques des populations microbiennes présentes dans l'environnement revêt aujourd'hui de nombreux intérêts non seulement au niveau fondamental mais aussi au niveau économique et industriel. Une meilleure caractérisation de l'ensemble des fonctions portées par ces microorganismes pourrait permettre d'une part, la découverte de nouvelles molécules à haute valeur ajoutée (arômes, antibiotiques, molécules anti-cancéreuses, etc.....) et d'autre part, le développement de procédés industriels plus efficaces et applicables par exemple en bioremédiation. Les biopuces ADN fonctionnelles (FGA) dédiées à l'étude des voies métaboliques de biodégradation des polluants sont particulièrement pertinentes pour répondre à cette problématique. En effet, par rapport à une approche de séquençage haut débit, elles ciblent uniquement les gènes responsables des métabolismes d'intérêt. Leur hybridation à la fois avec des cibles ADN et ARN extraits des mêmes échantillons environnementaux permet de percevoir très rapidement quels sont les gènes exprimés parmi le répertoire génique de la microflore présente. Cependant, pour que ces approches présentent un avantage indéniable, il faut qu'elles puissent donner accès à la fraction inconnue des populations microbiennes que l'on sait aujourd'hui considérable. Ceci est dorénavant possible grâce au développement de nouveaux logiciels qui autorisent le design de sondes dites exploratoires.

Avant nos travaux, le seul logiciel permettant de déterminer des sondes exploratoires dédiées à l'élaboration de FGA décrit dans la littérature était Metabolic Design (Terrat *et al.*, 2010 : ARTICLE ANNEXE 1). Son fonctionnement est basé sur l'utilisation d'alignements de séquences protéiques et sur l'identification de motifs conservés pour la sélection de sondes. La difficulté de cette approche est la capacité à déterminer des sondes longues (50 mers et plus) pour construire des FGA très sensibles même pour des cibles peu abondantes, ce qui est généralement le cas pour des gènes fonctionnels et leurs transcrits (Gentry *et al.*, 2006). La traduction inverse des motifs protéiques supérieurs à 7 ou 8 acides aminés génère le plus souvent des séquences nucléiques avec des taux de dégénérescence très élevés puisque l'ensemble de la dégénérescence du code génétique est prise en compte. Le nombre de sondes non dégénérées issu de ces séquences peut alors être considérable, et encore inadapté aux formats de biopuces disponibles actuellement sur le marché. En effet, si l'on considère que chaque acide aminé est codé en moyenne par trois codons différents (20 acides aminés spécifiés par 61 codons), chaque sonde dégénérée de 51 mers (soit l'équivalent de 17 acides aminés) produit plus de 3^{17} combinaisons non dégénérées (soit plus de 129 millions). De plus,

les tests de spécificité de l'ensemble de ces sondes nécessitent des temps de calcul très importants qui ne sont pas forcément à la portée des ressources informatiques présentes dans tous les laboratoires.

2. Objectifs

Afin d'apporter une nouvelle alternative à la conception de sondes longues, l'objectif de ce travail de recherche était de développer un nouveau logiciel offrant la possibilité de pouvoir sélectionner des sondes longues pour FGA sensibles, spécifiques et exploratoires en utilisant une interface web ne nécessitant donc pas, pour les utilisateurs, de posséder une puissance de calculs importante tout en s'affranchissant de la construction d'une base de données de référence pour les tests de spécificité. Un autre facteur important, pris en compte lors de sa conception, était la possibilité de le faire évoluer pour des applications concernant n'importe quelle problématique biologique (médicale, environnementale, agronomique, etc...). La stratégie employée est basée sur la sélection de sondes à partir (i) de séquences nucléiques correspondant à des gènes, ou (ii) de séquences consensus dégénérées, issues de l'alignement de toutes les séquences nucléiques disponibles pour une famille de gènes. Comparée à l'approche utilisant des séquences protéiques en entrée, cette stratégie génère des sondes beaucoup moins dégénérées ce qui autorise la sélection de sondes de taille supérieure. L'ensemble de la démarche mise en place a donné naissance au logiciel nommé HiSpOD pour : « High Specific Oligo Design ». Afin d'évaluer l'efficacité de ce nouvel outil, une FGA ciblant les principaux gènes impliqués dans la biodégradation des chloroéthènes a été construite. Elle a ensuite été hybridée avec des cibles simples composées d'ARN synthétiques correspondant à plusieurs gènes ciblés, et d'autres plus complexes constituées d'ARNm extraits d'une nappe phréatique polluée par du TCE en cours de traitement par biostimulation.

3. Principaux résultats

Le travail réalisé au cours de cette étude a permis de proposer un nouveau logiciel de design de sondes oligonucléotidiques utilisant une stratégie originale qui a donné lieu à une publication dans le journal « Bioinformatics ». Ce logiciel, nommé HiSpOD, est accessible au public par l'intermédiaire d'une interface web conviviale (<http://fc.isima.fr/~g2im/hispod/>) ce qui lui confère un avantage en terme de facilité d'utilisation et de diffusion. Il présente

également l'intérêt d'intégrer une base de données de CDS complète élaborée à partir des séquences comprises dans les divisions procaryote (PRO), champignon (FUN) et environnement (ENV) de la base de données de séquences nucléiques EMBL (<http://www.ebi.ac.uk/embl/>). L'utilisation de cette base est essentielle pour tester la spécificité des sondes sélectionnées. Par ailleurs, HiSpOD est le seul logiciel capable de prendre en compte à la fois tous les paramètres primordiaux en relation avec la sensibilité des sondes (taille, T_m , % GC, région de faible complexité) et l'utilisation de bases dégénérées nécessaire pour la création de sondes exploratoires.

La validation de la pertinence du logiciel HiSpOD a été effectuée grâce à la construction d'une FGA ciblant 21 gènes ou familles de gènes impliqués dans la biodégradation des chloroéthènes. Au total, 295 sondes de 50 mers ont été sélectionnées et utilisées pour la construction de la biopuce. Le temps nécessaire au logiciel pour le design des sondes est relativement court. Il faut compter en moyenne 3,5 h par gène lorsque la séquence donnée en entrée est une séquence unique, et 9,5 h pour une séquence consensus dégénérée. En général, cette stratégie permet la sélection de 8 à 9 sondes spécifiques pour chaque gène. L'utilisation de cibles de différente nature a permis d'évaluer la sensibilité et la spécificité de la FGA ainsi élaborée. Dans un premier temps, l'hybridation d'ARN obtenus à partir de gènes synthétiques correspondant aux gènes *mmoC* (méthane monooxygénase), *vcrA* (VC réductase) et *tceA* (TCE déhalogénase) a révélé un signal positif pour la totalité de leurs sondes spécifiques (n=52), et ceci avec une intensité suffisamment forte pour être facilement détectés (rapport signal/bruit de fond d'environ 1600 en moyenne), alors que l'ensemble des sondes ciblant les autres gènes présents sur la biopuce ne donnait aucune réponse positive (rapport signal/bruit < 5). Ces résultats traduisent, par conséquent, la grande spécificité des sondes sélectionnées et l'efficacité de notre outil bioinformatique.

Ensuite, afin de confirmer la puissance de notre FGA pour répondre à une question biologique, des ARNm ont été extraits d'un échantillon prélevé dans une nappe phréatique polluée par des chloroéthènes, puis utilisés comme cible. Leur hybridation a permis la détection de trois gènes, tous impliqués dans la voie métabolique de déchloration réductrice : le gène *pceA* à l'origine des premières étapes de dégradation des chloroéthènes et caractéristique du genre *Sulfurospirillum* ainsi que deux gènes responsables des étapes finales de la dégradation, *vcrA* et *bvcA*, et spécifiques des souches affiliées au genre *Dehalococcoides*. Comme les intensités de signal obtenues pour les sondes ciblant ces gènes sont comparables avec les niveaux d'expression déterminés par une approche qPCR à partir

du même échantillon, ces résultats soulignent, non seulement la puissance de cet outil en termes de spécificité et de sensibilité, mais également son intérêt d'un point de vue quantitatif.

HiSpOD: probe design for functional DNA microarrays

Eric Dugat-Bony^{1,2,†}, Mohieddine Missaoui^{3,4,†}, Eric Peyretailade^{2,5}, Corinne Biderre-Petit^{1,2}, Ourdia Bouzid^{1,2}, Christophe Gouinaud^{3,4}, David Hill^{3,4} and Pierre Peyret^{2,5,*}

¹Clermont Université, Université Blaise Pascal, Laboratoire Microorganismes: Génome et Environnement, BP 10448, F-63000, Clermont-Ferrand

²UMR CNRS 6023, Université Blaise Pascal, 63000 Clermont-Ferrand, France

³Clermont Université, Université d'Auvergne, Laboratoire Microorganismes: Génome et Environnement, BP 10448, F-63000, Clermont-Ferrand

⁴Clermont Université, Université Blaise Pascal, LIMOS, BP 10448, F-63000 Clermont-Ferrand

⁵UMR CNRS 6158, LIMOS, F-63173 Aubière.

Associate Editor: Trey Ideker. * To whom correspondence should be addressed. † The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

ABSTRACT

Motivation: The use of DNA microarrays allows the monitoring of the extreme microbial diversity encountered in complex samples like environmental ones as well as that of their functional capacities. However, no probe design software currently available is adapted to easily design efficient and explorative probes for functional gene arrays.

Results: We present a new efficient functional microarray probe design algorithm called HiSpOD (High Specific Oligo Design). This uses individual nucleic sequences or consensus sequences produced by multiple alignments to design highly specific probes. Indeed, to bypass crucial problem of cross-hybridizations, probe specificity is assessed by similarity search against a large formatted database dedicated to microbial communities containing about 10 million coding sequences (CDS). For experimental validation, a microarray targeting genes encoding enzymes involved in chlorinated solvent biodegradation was built. The results obtained from a contaminated environmental sample proved the specificity and the sensitivity of probes designed with the HiSpOD program.

Availability: <http://fc.isima.fr/~g2im/hispod/>.

Contact: pierre.peyret@univ-bpclermont.fr

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

Received September 27, 2010; Revision received November 23, 2010; Accepted December 15, 2010.

1 INTRODUCTION

Microbes mediate almost every conceivable biological process on the planet. It is estimated that about 10^{30} bacterial and archaeal cells are involved in these tremendous metabolic potentialities (Whitman *et al.*, 1998). To access to this wide metabolic diversity, environmental functional gene arrays (FGAs) are high-throughput tools which allow analyzing the presence or the expression of thousands of genes encoding key enzymes involved in many metabolic process in only one experiment. However, designing DNA microarray probes to encompass the full diversity of gene family sequences encountered in nature and not yet identified is still one of the most difficult challenges. In addition, the difficulty to obtain sufficient nucleic acid materials for microarray studies from environmental samples requires the development of highly sensitive probes. Currently, the most complete FGA available to monitor bacterial metabolism in complex ecosystems is GeoChip (He *et al.*, 2007) with a latest version covering ~47 000 gene variants from more than 290 gene categories (He *et al.*, 2008). GeoChip has already demonstrated its power in assessing *in situ* bacterial activity, reinforcing the idea that functional microarray are an efficient approach (Berthrong *et al.*, 2009; Kimes *et al.*, 2010; Leigh *et al.*, 2007; Liang *et al.*, 2009; Mason *et al.*, 2008; Rhee *et al.*, 2004; Van Nostrand *et al.*, 2009; Wang *et al.*, 2009; Yergeau *et al.*, 2007). However, the main problem that must be faced for functional microarray applications is the complexity of the oligonucleotide design step which hinders their development.

Many parameters such as probe length, melting temperature (T_m), %GC, complexity or still potential cross-hybridizations evaluation (percent similarity and identity stretch between probe and non-targeted sequences) must be taken into account to ensure the selection of the oligonucleotides offering the best compromise between specificity and sensitivity (Kane *et al.*, 2000; Pozhitkov *et al.*, 2007). Fortunately, numerous bioinformatics tools, which estimate most of these parameters, have been developed to perform this crucial probe design step (Lemoine *et al.*, 2009). Currently, the most popular software for FGA probe design are OligoWiz (Wernersson and Nielsen, 2005), CommOligo (also named OligoStar) (Li *et al.*, 2005) or YODA (Nordberg, 2005). However, these softwares were developed for use on whole-genome data or on specific sets of genes (Gentry *et al.*, 2006; Lemoine *et al.*, 2009). So, cross-hybridization research is performed only against a reduced set of sequences (Li *et al.*, 2005; Wang and Seed, 2003; Wernersson and Nielsen, 2005). Thus, probe specificity must be checked against databases containing a very large set of sequences while staying within a reasonable time for the design.

A second critical obstacle when dealing with environmental studies is the difficulty to design probes able to cover all the gene variants encoding proteins having a same function (Habe and Omori, 2003). Although Hierarchical Probe Design (Chung *et al.*, 2005) and ProDesign (Feng and Tillier, 2007) softwares propose clustering methods to identify conserved regions allowing a probe design for gene families, they are unable to perform the specificity tests. The single promising approach to address this problem is implemented in the PhylArray program (Milton *et al.*, 2007) which consists in the design of degenerate probes targeting multiple sequence variants. This strategy allows the design of probes targeting sequences never described in international databases and therefore considered as explorative probes. Unfortunately, this tool is only dedicated to probe design for phylogenetic microarrays.

Finally, despite the wide range of available programs, few provide access to oligonucleotide design through a web interface. This solution is the easiest one since it does not require any local software installation. Unfortunately, only six softwares offer this possibility and among them four (OligoFaktory, Osprey, PROBEmer and ROSO) allow the user to implement its own biological models for a tailor-made probe design.

Therefore, many improvements in the probe design technology are still necessary and the HiSpOD (High Specific Oligo Design) program was developed to meet these challenges. In addition to take into account the crucial parameters for the design of effective probes, we combine:

- the possibility to produce specific probes and also degenerate probes from consensus sequences, and thus with an explorative power, in order to detect a broad spectrum of variants from gene families;
- the use of a large formatted database dedicated to microbial communities [all coding DNA sequences (CDSs) from Prokaryotes (PRO), Fungi (FUN) and Environmental (ENV) taxonomic divisions of the EMBL databank] to check specificity of selected probes and limit cross-hybridizations;
- program parallelization in order to optimize the speed of the design;
- the oligonucleotide design access through a web interface.

This user-friendly bioinformatics tool is dedicated to probe design for studying microbial communities. It can also be applied for all research areas after adaptation of the database used

to search for cross-hybridizations. The algorithm and its features are discussed and demonstrated with the analyses of a microarray built from gene sequences encoding enzymes involved in chlorinated solvent [tetrachloroethene (PCE) and trichloroethene (TCE)] biodegradation. The functional DNA microarrays developed in this study will be particularly useful to monitor *in situ* bacterial capacities on contaminated environments.

2 COMPLETE CDS DATABASE BUILDING

A new database was developed in order to assess the potential cross-hybridizations of probes designed with HiSpOD by using non-target sequences present in environmental samples. All annotated transcript sequences in all classes of EMBL databank release 97 of PRO, FUN and ENV taxonomic divisions (<ftp.ebi.ac.uk/pub/databases/embl/release>) were extracted using a Perl script named '*buildDatabase.pl*'. Briefly, the script downloads compressed files containing EMBL entries from the EMBL databank, extracts the CDS from each EMBL entry and their associated 5' and 3' untranslated regions (UTR) fixed by default as the 300 upstream and downstream bases to the CDS. Indeed, UTRs were considered to avoid cross-hybridizations because they are components of mRNA sequences. To obtain a suitable and curated database, a filtering step is carried out to ensure that sequences with bad quality were rejected if:

1. The percentage of *N* (unknown base) in the sequence is >10%.
2. At least one stretch of 10 consecutive *N* is detected.

The data extracted are then saved into a relational database implemented under Oracle and named EnvExBase. Finally, the whole database contains 9 129 323 sequences (about 10 GB). Proteins obtained from each original CDS were also saved in a separate table in order to allow rapid accession of their specific information like source organisms and biological functions of encoded proteins.

3 DESIGN ALGORITHM

HiSpOD software allows oligonucleotide probe design for FGAs in the context of microbial ecology but could also be adapted to various biological models. It needs an input file in FASTA format containing the target sequences. Two kinds of sequence are accepted by the program: specific nucleic acid sequence and degenerate nucleic acid sequence [consensus sequence resulting from multiple alignments and based on IUB/IUPAC code (International

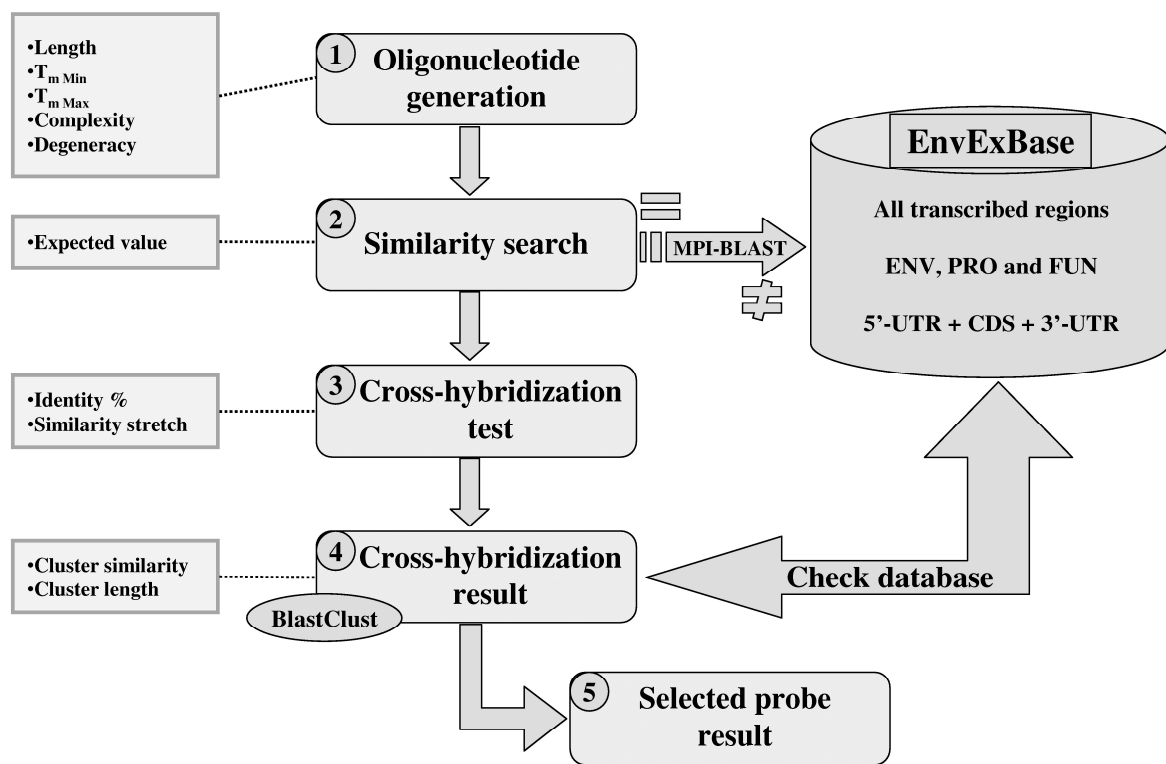


Fig. 1. HiSpOD design workflow. HiSpOD program is composed of five steps and each parameter chosen by the user are indicated on the left.

Union of Biochemistry and Molecular Biology/International Union of Pure and Applied Chemistry)]. The program workflow consists in five consecutive steps (Fig. 1):

1. Oligonucleotide probe design. For each input sequence, the probe length is defined by the user. Along the sequence, the algorithm extracts all probes by incrementing the constant defined probe size in a window. First, in the case of degenerated sequences, all potential oligonucleotides generated with a maximal degeneracy higher than the threshold fixed by the user are removed. Then, all oligonucleotides or all combinations resulting from degenerate probes are filtered according to the melting temperature (minimum and maximum T_m determined by the formula (A) in which $[Na^+]$ is fixed by default at 0.5 M) and the complexity stretch (maximum authorized number of identical consecutive bases).

$$T_m = 79.8 + 18.5 \times \log_{10}([Na^+]) + 58.4 \times \frac{yG + zC}{wA + xT + yG + zC} + 11.8 \times \frac{(yG + zC)^2}{(wA + xT + yG + zC)^2} - \frac{920}{wA + xT + yG + zC} \quad (A)$$

2. Similarity search. Oligonucleotides which successfully pass the first step are tested for similarity search against all the sequences from EnvExBase database using BLASTN program (Altschul *et al.*, 1990). The BLAST program is tuned for this task as follows: the 'DUST' filter, used to mask low complexity regions from nucleic acid sequences, is inactive (-F F), the search strand is set to the positive one (-S 1), the expected-value (-e) is a parameter chosen by the user (1000 by default), the word size is set to the minimum length (-W 7) and the number of reported results is fixed to 500 (-b 500).
3. Cross-hybridization test. Kane's criteria (Kane *et al.*, 2000), e.g. percent similarity and identity stretch, defined by the user (default values: 85 and 15, respectively), are checked for all positive results from previous BLASTN analysis to identify potential cross-hybridization.
4. Cross-hybridization result. All sequences identified to potentially cross-hybridize with designed probes are clustered using BLASTClust tool provided by the BLAST program. The options (-S) which gives the minimum percentage of similarity between sequences that should be grouped together and (-L) which gives the minimum percentage of length of the implicated sequences in the alignment that should be considered, are also fixed by the user (default values: 90 and 0, respectively).

5. Selected probe result. The final result provided by our software is stored within two output files: the first file with the extension '.probes' included all filtered oligonucleotides sorted by the number of cross-hybridizations. The position on the input sequence, the length and the sequence of the oligonucleotide are also mentioned. The second file '.result' details the non-targeted sequences that may induce potential cross-hybridizations with their accession numbers, source organism and function of the encoded protein.

4 IMPLEMENTATION

HiSpOD is written in Perl 5 scripting language and uses a BioPerl module for better visibility and to simplify the source code. A parallel implementation allows probe finding to be done in parallel on computing cluster architecture. The parallelism is introduced in Step (3) (cross-hybridization test) which is the most time-consuming task of the process. For this step, HiSpOD integrates a distributed and parallel implementation of the BLAST program (Darling *et al.*, 2003) which uses the MPI library (Message Passing Interface) and runs on parallel architectures.

HiSpOD software is installed on a symmetric multiprocessor (SMP) machine with 8 AMD Opteron Dual-Core processors at 1.8 GHz, 128 GB RAM and running Linux CentOS 64 bits. For large data storage, a SAN system disk with 2TB is used.

The application is freely accessible from a web interface developed in PHP/MySQL at <http://fc.isima.fr/~g2im/hispod/>. The design results are made available on the server. HiSpOD is started via the interface using the chosen data entered by the user. The computing is made on the symmetric multiprocessor (SMP) machine after a remote connection.

5 EXPERIMENTAL VALIDATION

The efficiency of HiSpOD was then assessed by the design of probes targeting genes encoding enzymes involved in the chlorinated solvent biodegradation.

5.1 Oligonucleotide probe set design

Probes were designed from 54 nucleic sequences corresponding to 21 genes coding for enzymes involved in chlorinated solvent biodegradation pathways (15 involved in anaerobic pathways and 6 in aerobic ones), which were downloaded from the NCBI database

Table 1. Input parameters for HiSpOD probe design

Probe characteristics				Probe specificity			
Probe length (mers)	T_m range (°C)	Single nucleotide stretch length (complexity)	Maximal degeneracy	Maximal consecutive match with non-target	Maximal identity with non-target (%)	BLASTClust option (-S) (%)	BLASTClust option (-L) (%)
50	64–79	10	1 or 32	15	75	90	0

(-L)=0 indicates that sequence length is not considered for the clustering.

Table 2. Total number of selected probes and time consuming for the design for individual gene sequences and degenerate consensus sequences (gene family)

Data observed	Individual gene sequences	Degenerate consensus sequences
Design time (min)	3256	4014
Average design time per sequence (min)	203.5	573
No. of probes selected	140	54 (155 ^a)
Average No. of probes per sequence	9	8 (22 ^a)

^aNumber of probes after non-degenerate probe creation from each degenerate ones.
No. number.

(Supplementary Table S1). Fourteen genes were treated, starting with individual nucleotide sequences (*amoB*, *todC1*, *todC2*, *bmoC*, *mmoC*, *xamoA*, *bvcA*, *prdA*, *rdhA1*, *rdhA2*, *pceA* from *Dehalococcoides ethenogenes* 195, *pceA* from *Suflurospirillum multivorans*, *pceA* from *Dehalobacter restrictus* PER-K23 and *pceA* from *Shewanella sediminis* HAW-EB3), 5 genes were treated at the gene family level (*pceA* from *Geobacter*, *rdhA1A*, *rdhA1B*, *rdhA2A* and *rdhA2E*) and 2 were treated at both the individual and the family levels (*tceA* and *vcrA*). In order to target gene families, multiple sequences were aligned by using ClustalW program (Thompson *et al.*, 1994) to determine degenerate consensus sequences before probe design with HiSpOD. For each job launched on HiSpOD program (23 entries), probes of 50mers in length were designed. Table 1 summarizes the parameter values used for the probe design.

T_m range was determined to obtain a probe set with a GC content between 40 and 60%. Thus, ~5 days (121 h) were necessary for the complete design with an average of 3.5 h for an individual sequence and of 9.5 h for a degenerate sequence. Based on the specificity (no or few cross-hybridizations) and the location of the probe along the target sequence, multiple probes were then selected per gene (Table 2 and Supplementary Fig. S1).

In total, 295 oligonucleotide probes were designed. These were used to elaborate a functional DNA microarray.

5.2 Experimental procedures

5.2.1 Microarray building

Microarrays were produced with the NimbleGen high-density array synthesis technology (Roche NimbleGen, Madison, Wisconsin, USA). Each oligonucleotide was synthesized *in situ* in triplicate. Probes were randomly distributed across the array in order to minimize spatial effect. In addition, the microarray also reported 8863 random probes (20–56mers) used as a metric of technical background noise and 11 573 other control probes (positive and negative) which allowed quality control of oligonucleotide synthesis and hybridization conditions.

5.2.2 Synthetic antisense RNA target preparation and labelling

The complete *mmoC* (S81887), *vcrA* (AY322364) and *tceA* (AF228507) gene sequences associated to SP6 promoter sequence at their 3' end were synthesized by Biomartik Corporation (Cambridge, Ontario, Canada) and cloned into the pGH vector. Antisense RNAs (aRNAs) were produced by *in vitro* transcription using the MEGAscript SP6 Kit (Ambion,

Austin, Texas, USA) and then labelled with Cy3-ULS as described in the Kreatech ULS labelling procedure (Kreatech Diagnostics, Amsterdam, The Netherlands). The aRNAs were quantified by spectrophotometry (at 260 nm) as well as dye incorporation (at 550 nm) with a Nanodrop ND-1000 (NanoDrop Technologies, Wilmington, North Carolina, USA) and RNA integrity was controlled with a Model 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, California, USA).

5.2.3 Environmental target preparation and labelling

Total RNAs were extracted according a modified protocol (Vetriani *et al.*, 2003) from an industrial site contaminated with both tetrachloroethene (PCE) and TCE) and biostimulated during 31 months before sample collection by SITA Remediation Company (Lyon, France). In this protocol, freeze-thaw cycles were replaced by beat betting 1 min at 30 Hz with 1 g of glass beads of $\leq 106\mu\text{m}$ in diameter (Sigma-Aldrich, Saint Louis, Missouri, USA). Co-extracted DNA was removed by digestion with 4U of DNase I (DNA-free, Ambion, Austin, Texas, USA) at 37°C for 35 min. Using the MicroExpress kit (Ambion, Austin, Texas, USA), 8–9 μg of total RNA mixture were subjected to mRNA enrichment protocol. In a second step, antisense amino-allyl dUTP marked RNA (aRNA) was obtained by amplification with the MessageAmp II-Bacteria kit (Ambion, Austin, Texas, USA) and labelled with Cy3 fluorescent dye (GE Healthcare, Chalfont St Giles, UK) following the manufacturer's instructions. RNA integrity was controlled with a Model 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, California, USA).

5.2.4 Microarray hybridization

A microarray was hybridized with 10^{11} copies of each labelled Cy3-aRNA obtained from synthetic genes and another one with 6 μg of Cy3-aRNA obtained from the contaminated groundwater. Targets were dried in a SpeedVac (Thermo Fisher Scientific, Villebon sur Yvette, France) and resuspended in the NimbleGen Hybridization Buffer (Roche NimbleGen, Madison, Wisconsin, USA). Then, targets were hybridized on the microarray at 42°C for 72 h using a 4-bay NimbleGen Hybridization system (Roche NimbleGen, Madison, Wisconsin, USA). Array washes were performed as recommended by NimbleGen and slides were scanned at 2 μm resolution using the InnoScan[®] 900 and Mapix[®] software (Innopsys, Carbonne, France). Finally, pixel intensities were extracted using NimbleScan[™] v2.5 software (Roche NimbleGen, Madison, Wisconsin, USA) and pair reports containing signal intensity data for every spot on the array linked to its corresponding probe identifier were generated.

5.2.5 Data normalization and statistical analysis

The background noise was determined using random probes present on the microarrays with the method described in Supplementary Material 1 and was defined by two parameters: the background median intensity (B_{position}) and its dispersion ($B_{\text{dispersion}}$). Finally, a modified signal-to-noise ratio named SNR' and based on the formula $\text{SNR}' = (\text{probe signal intensity} - B_{\text{position}})/B_{\text{dispersion}}$ was calculated in order to normalize our data. As suggested in the study of He and Zhou (2008), positive hybridization was considered significant for probes having a $\text{SNR}' > 5$.

5.2.6 Gene isolation from the contaminated groundwater

Genomic DNA (gDNA) was extracted from the industrial contaminated site using a modified protocol originally described by Vetriani *et al.* (2003). In this protocol, freeze-thaw cycles were replaced by beat betting 1 min at 30 Hz with 1 g of glass beads of $\leq 106\mu\text{m}$ in diameter (Sigma-Aldrich, Saint Louis, Missouri, USA). gDNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, North Carolina, USA). PCR reactions were carried out to isolate gene fragments for *vcrA*, *bvcA*, *pceA* and *todC1* (primer list and PCR conditions in Supplementary Table S2). Amplicons of correct size were then cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, California, USA) and sequenced by MWG DNA sequencing services (Ebersberg, Germany).

5.2.7 Quantitative RT-PCR

In order to validate the semi-quantitative aspect of microarray results, qRT-PCR assays were performed. RT reactions were carried out in duplicate from 100 ng of total RNA extracted from the contaminated environment using *vcrA*, *bvcA* and *pceA* (0.625 μM of each primer) reverse primer mix (Supplementary Table S2). Control for DNA contamination evaluation was the same reaction without reverse transcriptase. Then, quantitative reactions were performed for each gene independently. For each of the two RT replicates, the quantitative reaction was achieved twice, thus leading to four measurements for each gene. For the gDNA contamination control, the quantitative reaction was completed four times. Finally, three replicates for each point of the standard curve (serially diluted cDNA) were measured. The reaction was performed in a final volume of 20 μl containing 5 μl of cDNA, 10 μl of 2X MESA Green qPCR for SYBR assay mixture (Eurogentec, Liege, Belgium) and the corresponding primer sets described in Supplementary Material (Supplementary Table S3) at

Table 3. Probe sensitivity and specificity assessment after microarray hybridization with three aRNA targets (*mmoC*, *vcrA* and *tceA*)

Targeted gene	No. of positive probes/No. of total probes	Average SNR'
<i>mmoC</i>	5/5	1522.5
<i>vcrA</i>	38/38	1624.5
<i>tceA</i>	9/9	1709.4
Others	0/243	< 5

0.2µM final concentration each. qPCR was realized in the Mastercycler Realplex (Eppendorf, Hamburg, Germany) starting with 5 min of denaturation at 95°C followed by 40 cycles consisting of denaturation at 95°C for 15 s, annealing at 59°C for 15 s and elongation at 68°C for 30 s. Data analysis was achieved with *realplex* software version 1.5 (Eppendorf, Hamburg, Germany). The quantification of transcripts in the RNA sample was determined after subtraction of the gene copy number obtained for the gDNA contamination control.

5.2.8 Accession numbers

Nucleotide sequences without PCR primer sequences were deposited in the GenBank database under accession numbers HM140426 and HM140427 for *vcrA*, HM140428 and HM140429 for *bvcA* and HM140430 and HM140431 for *pceA*. The microarray data discussed in this publication are available at the GEO web site (<http://www.ncbi.nlm.nih.gov/gate1.inist.fr/geo/>) under accession number: GSE21492.

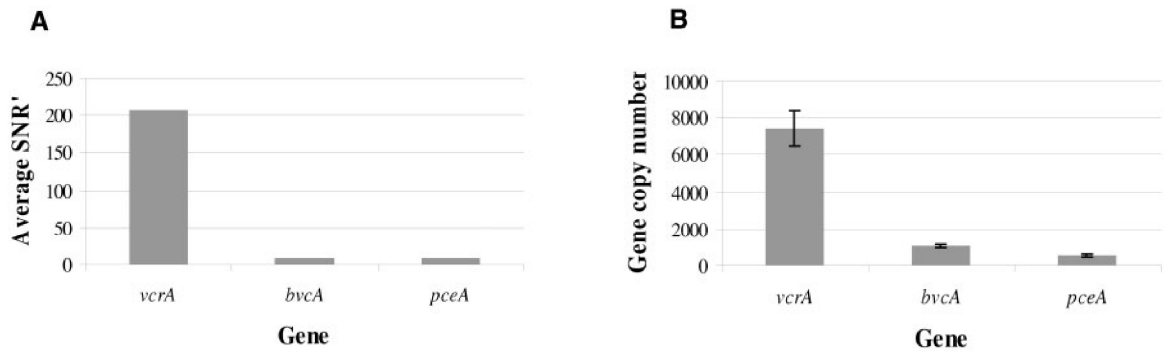
5.3 Probe efficiency analysis

5.3.1 Probe sensitivity and specificity

A first experiment was performed to test the ability of probes to reliably identify genes involved in chlorinated solvent degradation. Thus, three synthetic genes [*mmoC* (S81887), *vcrA* (AY322364) and *tceA* (AF228507)] were used to produce aRNA by *in vitro* transcription. Then, 10¹¹ copies of each labelled aRNA were hybridized to the array. Whatever the gene tested, the results (GSE21492) showed that all corresponding probes gave a strong positive hybridization signal (Table 3), which is in agreement with a great sensitivity of probes designed with HiSpOD. These probes are also highly specific because no hybridization signals (SNR' > 5) were obtained with probes targeting the other genes (21 genes are targeted by the microarray).

5.3.2 Industrial polluted site analysis

Microarray hybridization with labelled antisense mRNA extracted from a PCE/TCE contaminated groundwater, allowed a full identification of expressed genes on the environmental sample. Four genes were detected (average intensity of all specific probes with SNR' > 5): three encoded reductive dehalogenases involved in the anaerobic dehalorespiration pathway (*vcrA* and *bvcA* from *Dehalococcoides* and *pceA* from *Sulfurospirillum*) and one implicated in the aerobic TCE biodegradation pathway (*todC1* from *Pseudomonas*). Positive

**Table 4.** Comparison of parameters used by HiSpOD and by three popular softwares

hybridization signals were obtained for all regions of the three genes involved in the anaerobic dehalorespiration pathway. In contrast, only one out of the two targeted regions for *todC1* was detected. This finding suggested that hybridization of the probes designed to target the 3' extremity of *todC1* gene resulted from cross-hybridization and did not reflect the presence of transcripts on the contaminated site. In addition, PCR and sequencing experiments allowed the isolation of gene fragments for *vcrA* (HM140426 and HM140427), *bvcA* (HM140428 and HM140429) and *pceA* (HM140430 and HM140431) but not for *todC1*, confirming the results obtained with the DNA microarray.

5.3.3 Gene expression quantification

SNR' levels obtained for *vcrA* gene (average SNR' = 205.6 for the individual gene design and average SNR' = 220.6 for the gene family design) were much higher than those obtained for the last two genes *bvcA* and *pceA* (average SNR' = 7.3 and 8.2, respectively), which could suggest a higher abundance of *vcrA* transcripts in the sample (Fig. 2A). To confirm these results, transcript quantification by qRT-PCR was performed on the same groundwater sample and the measurement of the expression level showed a higher number of transcripts for *vcrA* (7415 ± 949 copies/ng of RNA) than for both *bvcA* and *pceA* (1072 ± 86 and 559 ± 87 copies/ng of RNA, respectively) in the RNA sample (Fig. 2B).

These data support the applicability of our FGA for gene identification and as a semi-quantitative tool for gene expression evaluation with the advantage to survey numerous genes at the same time.

6 DISCUSSION

HiSpOD takes into account a greater number of criteria to select more efficient probes (Table 4) compared with most popular softwares used for oligonucleotide probe design for FGAs (Li *et al.*, 2005; Nordberg, 2005; Wernersson and Nielsen, 2005). The entire software suite is available through web services. To produce the DNA probes, users can enter their sequences of interest and set the required parameters.

Furthermore, our software can be applied to both, individual nucleotide gene sequences and degenerated consensus sequences, produced by multiple alignments of gene sequences belonging to a same gene family. Therefore, the user has the possibility to identify specific probes either at the gene or at the gene family level. For this last case, design methods

previously reported are based on sequence clustering (Chung *et al.*, 2005; Feng and Tillier, 2007) or on the use of mismatch probes (Jaing *et al.*, 2008), which constrain the selection of probes in the most conserved regions. In contrast, HiSpOD allows considering regions generating the least number of cross-hybridizations even if they are more variable within the gene family. Because all sequence combinations for each degenerate probe are re-created for the DNA microarray synthesis, some of them could detect new gene sequences not yet available in international databases. Furthermore, as the number of probes per array is no longer a constraint, given the development of very high density microarrays (several million probes proposed by NimbleGen), designing microarray to encompass the full diversity of sequence variants encountered in the environment or in specific metabolic process is now feasible. By this aspect, the HiSpOD method is then a pioneer explorative approach to the design of probes dedicated to FGAs. Today, the only probe design software described to allow an exploratory study of environmental samples using degenerate probes is PhylArray (Milton *et al.*, 2007), but it is dedicated to phylogenetic oligonucleotide microarrays (POAs).

Cross-hybridizations with non-target sequences can be a significant contribution to the hybridization signal, potentially introducing substantial error. According to He *et al.* (2007), they represent one of the most critical limits of DNA microarray approach when dealing with complex environmental samples of unknown composition. However, over the past two decades, new sequencing technology developments and the trend of declining sequencing costs, allowed an exponential growth of public DNA sequence data. Data analysis has revealed an extraordinary gene diversity and a huge nucleic sequence polymorphism for all investigated environmental microbial habitats especially by metagenomics approach (Kottmann *et al.*, 2010). This extraordinary resource can now be mined for DNA microarrays applications to assess the potential cross-hybridizations. Currently, HiSpOD is the most suitable tool to take into account this wide diversity to limit cross-hybridizations with a test of specificity conducted against EnvExBase, a large and representative database of all prokaryotic, fungal and environmental sequences. This database includes all known potential CDSs and their putative UTRs. However, as the HiSpOD algorithm extracts all probes by incrementing the constant defined probe size in a window along the sequence, the time required for evaluating their specificity can be quite long (several hours per gene) especially for highly degenerate oligonucleotide probes. So, HiSpOD is greatly dependent on the performance of MPI-BLAST program used in the specificity test step. Currently, we are focusing our efforts to increase the program performance by deploying it on a larger cluster

infrastructure and on a computing grid. Given the exponential increase in the number of sequences in databases, the DNA microarrays will be upgraded regularly with novel probes allowing the targeting of either new variants of the targeted genes or genes newly discovered. In addition, EnvExBase component of the HiSpOD software will be updated regularly. An improvement of the database structure, consisting in a split into multiple sub-databases according to particular ecosystems (soil, marine, freshwater, atmosphere, etc.), is also under development. It will provide an enhanced usability of the program, with a faster speed and an elimination of unneeded cross-hybridization information. Currently, no known software offers the opportunity to test for specificity on such a large database. Furthermore, no tested software (i.e. OligoWiz, CommOligo and Yoda) is able to integrate this database to perform the specificity test with it. This is probably due to an inadequate performance optimization of these algorithms which systematically crashed when using a large database. In addition, the clustering of cross-hybridization sequences using BLASTClust tool is another original approach proposed by our software to define gene sequence families that may induce potential cross-hybridizations and so to facilitate selection of the most appropriate probes.

Probe sensitivity is another crucial parameter, particularly for environmental studies, where biomass can be low. So, design process allowing the selection of oligonucleotide probes need to be optimized. Generally, microarray-based hybridization presents a low detection limit. For example, 50mer probe FGA detection limit was estimated to be in the range of 5–10 ng of gDNA in the absence of background DNA and was about 10-fold lower in its presence (Rhee *et al.*, 2004). In this last case, target sequences were extrapolated to represent genomic material from <5% of the total community of an environment (Gentry *et al.*, 2006). For mRNA, no detection limit is currently assessed in environmental sample. It remains hard to solve the sensitivity issue despite the existence of some protocols bringing those limits down (Gao *et al.*, 2007; Wu *et al.*, 2006). In this study, we demonstrated the ability of probes designed with HiSpOD to detect mRNA retrieved from environmental samples with a high sensitivity (e.g. 559 ± 87 copies/ng of RNA for *pceA*) (Fig. 2). Furthermore, expression profiles obtained were comparable with those of qRT-PCR assays suggesting the semi-quantitative aspect of the microarray elaborated using this software. We demonstrate that the microarray developed in this study provides a powerful tool to monitor chlorinated ethene biodegradation capabilities in complex environments. This tool will be now used for a wider study incorporating several industrial polluted sites where a more or less complete degradation of chlorinated solvents was observed.

However, certain limitations still remain. Indeed, numerous authors have highlighted the problem of signal variability obtained with probes targeting different regions of the same gene (Bruun *et al.*, 2007; Held *et al.*, 2006), certainly when linked with probe thermodynamic properties and/or target secondary structure. The function of these parameters is still difficult to evaluate and to visualize in the microarray results. Although their calculation is time consuming, they will be taken into account in the next version of HiSpOD.

In this study, the new software was shown to be particularly accurate to design probes dedicated to microbial ecology. The experimental validation of the oligonucleotide probe set designated using HiSpOD was performed with mRNA extracted from a PCE/TCE contaminated aquifer where pollutant biodegradation was observed (Site description and chemical analysis in Supplementary Material 2 and Table S4). These pollutants undergo biodegradation through two different pathways: (i) use as an electron acceptor (reductive dechlorination) (Futagami *et al.*, 2008) and (ii) co-metabolism where degradation of the chlorinated solvent is simply fortuitous and provides no benefit to the microorganism (Arp *et al.*, 2001), the first being the major process for the natural biodegradation of chlorinated solvents. In this study, the most recognized gene sequences involved in these two metabolic processes were extracted from international databases and were used for probe design. The microarray presented here is, to our knowledge the most complete high-throughput molecular tool describing these pathways with 295 probes covering 54 sequence variants of 21 genes and in addition with an explorative side. By comparison, GeoChip mentioned 35 probes targeting genes characteristics of PCE/TCE biodegradation pathways (He *et al.*, 2007). In the context of industrial site rehabilitation, it could be also applied as a diagnostic tool to detect microbial activities before, during and after bioremediation treatment. Moreover, this explorative high-throughput tool could be used to monitor the distribution of these catabolic pathways along non-contaminated ecosystems in order to better understand the origin of these fascinating microbial functions.

7 CONCLUSION

In summary, we present a novel probe design software called HiSpOD dedicated to FGA developments. HiSpOD allows probe design at both specific gene and gene family levels through an original approach based on degenerated probe determination. Moreover, specificity test was conducted against a large formatted database composed of all known CDS retrieved from the taxonomic divisions PRO, FUN and ENV of the EMBL databank in order

to avoid cross-hybridization events. Finally, the user-friendly web interface developed simplifies greatly the user access to the program.

ACKNOWLEDGEMENTS

We are grateful to our undergraduate bioinformatics students Nicolas Parisot (IUT Génie Biologique option bio-informatique, Aurillac) and Mathieu Roudel (Master Bioinformatique, Université Blaise Pascal, Clermont-Ferrand) for the development of statistical and annotation scripts used in HiSpOD, Pascale Gouinaud for her assistance and Dr Biron David for reviewing the English version of the manuscript. We thank SITA Remediation society for sample collection.

Funding: ‘Agence De l'Environnement et de la Maîtrise de l'Energie’ (ADEME, France) (grant ID 2598); INSTRUIRE, PREVOIR and LifeGrid programs financed by the regional council of Auvergne (France), the FEDER European commission and the ‘Centre national de la Recherche Scientifique’ (CNRS, France); the grant ANR-07-ECOT-005-05 for the program PRECodd Evasol from ‘Agence Nationale de la Recherche’ (ANR, France).

Conflict of Interest: none declared.

REFERENCES

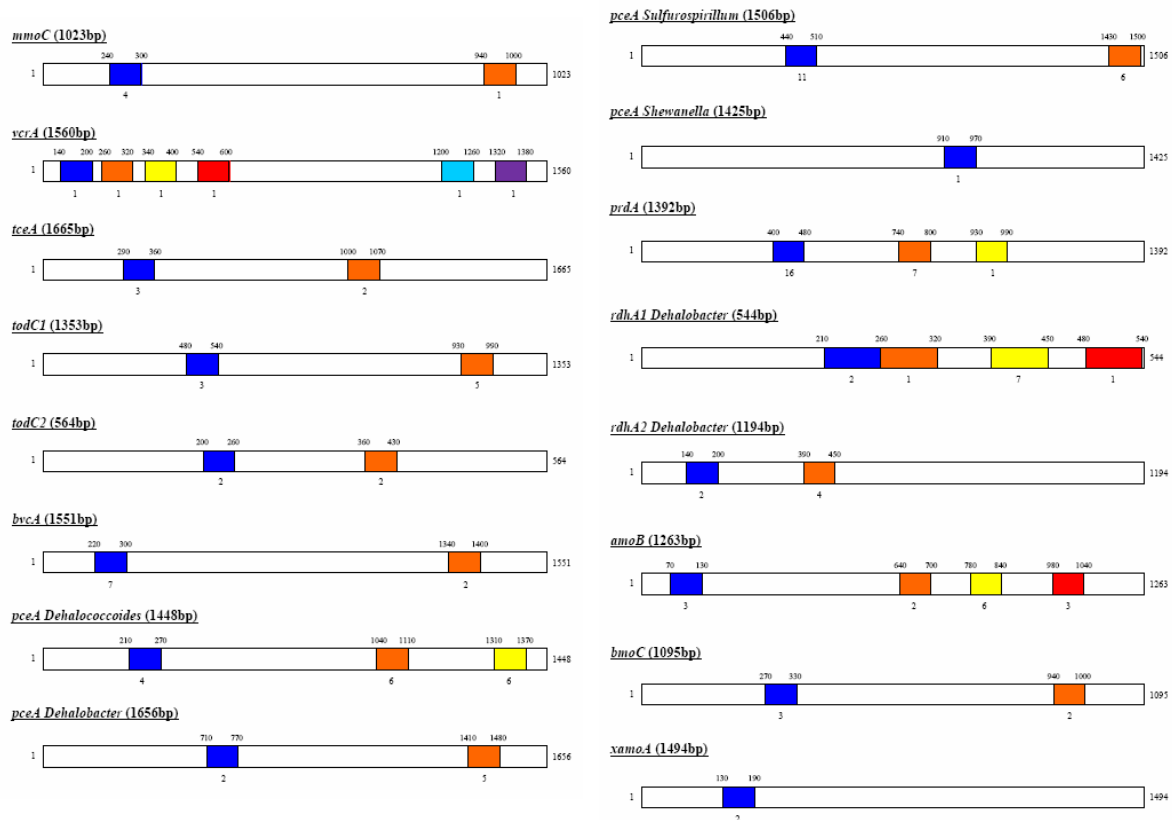
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool, *J. Mol. Biol.*, **215**, 403-410.
- Arp, D.J., Yeager, C.M. and Hyman, M.R. (2001) Molecular and cellular fundamentals of aerobic cometabolism of trichloroethylene, *Biodegradation*, **12**, 81-103.
- Berthrong, S.T., Schadt, C.W., Pineiro, G. and Jackson, R.B. (2009) Afforestation alters soil functional gene composition and biogeochemical processes in South American grasslands, *Appl. Environ. Microbiol.*, AEM.01126-01109.
- Bruun, G.M., Wernersson, R., Juncker, A.S., Willenbrock, H. and Nielsen, H.B. (2007) Improving comparability between microarray probe signals by thermodynamic intensity correction, *Nucleic Acids Res.*, **35**, e48.
- Chung, W.-H., Rhee, S.-K., Wan, X.-F., Bae, J.-W., Quan, Z.-X. and Park, Y.-H. (2005) Design of long oligonucleotide probes for functional gene detection in a microbial community, *Bioinformatics*, **21**, 4092-4100.
- Darling, A., Carey, L. and Feng, W. (2003) The Design, Implementation, and Evaluation of mpiBLAST, *4th International Conference on Linux Clusters: The HPC Revolution 2003*.
- Feng, S. and Tillier, E.R.M. (2007) A fast and flexible approach to oligonucleotide probe design for genomes and gene families, *Bioinformatics*, **23**, 1195-1202.
- Futagami, T., Goto, M. and Furukawa, K. (2008) Biochemical and genetic bases of dehalorespiration, *Chem. Rec.*, **8**, 1-12.

- Gao, H., Yang, Z.K., Gentry, T.J., Wu, L., Schadt, C.W. and Zhou, J. (2007) Microarray-Based Analysis of Microbial Community RNAs by Whole-Community RNA Amplification, *Appl. Environ. Microbiol.*, **73**, 563-571.
- Gentry, T., Wickham, G., Schadt, C., He, Z. and Zhou, J. (2006) Microarray Applications in Microbial Ecology Research, *Microb. Ecol.*, **52**, 159-175.
- Habe, H. and Omori, T. (2003) Genetics of Polycyclic Aromatic Hydrocarbon Metabolism in Diverse Aerobic Bacteria, *Biosci. Biotechnol. Biochem.*, **67**, 225-243.
- He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., Wu, W., Gu, B., Jardine, P., Criddle, C. and Zhou, J. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes, *ISME J.*, **1**, 67-77.
- He, Z. and Zhou, J. (2008) Empirical Evaluation of a New Method for Calculating Signal-to-Noise Ratio for Microarray Data Analysis, *Appl. Environ. Microbiol.*, **74**, 2957-2966.
- He, Z.L., Van Nostrand, J.D., Wu, L.Y. and Zhou, J.Z. (2008) Development and application of functional gene arrays for microbial community analysis, *Trans. Nonferrous Met. Soc. China*, **18**, 1319-1327.
- Held, G.A., Grinstein, G. and Tu, Y. (2006) Relationship between gene expression and observed intensities in DNA microarrays--a modeling study, *Nucleic Acids Res.*, **34**, e70.
- Jaing, C., Gardner, S., McLoughlin, K., Mulakken, N., Alegria-Hartman, M., Banda, P., Williams, P., Gu, P., Wagner, M., Manohar, C. and Slezak, T. (2008) A Functional Gene Array for Detection of Bacterial Virulence Elements, *PLoS ONE*, **3**, e2163.
- Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D. and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays, *Nucleic Acids Res.*, **28**, 4552-4557.
- Kimes, N.E., Nostrand, J.D.V., Weil, E., Zhou, J. and Morris, P.J. (2010) Microbial functional structure of "Montastraea faveolata", an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies, *Environ. Microbiol.*, **12**, 541-556.
- Kottmann, R., Kostadinov, I., Duhaime, M.B., Buttigieg, P.L., Yilmaz, P., Hankeln, W., Waldmann, J. and Glockner, F.O. (2010) Megx.net: integrated database resource for marine ecological genomics, *Nucleic Acids Res.*, **38**, D391-395.
- Leigh, M.B., Pellizari, V.H., Uhlik, O., Sutka, R., Rodrigues, J., Ostrom, N.E., Zhou, J. and Tiedje, J.M. (2007) Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs), *ISME J.*, **1**, 134-148.
- Lemoine, S., Combes, F. and Le Crom, S. (2009) An evaluation of custom microarray applications: the oligonucleotide design challenge, *Nucleic Acids Res.*, **37**, 1726-1739.
- Li, X., He, Z. and Zhou, J. (2005) Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation, *Nucleic Acids Res.*, **33**, 6114-6123.
- Liang, Y., Li, G., Nostrand, J.D.V., He, Z., Wu, L., Deng, Y., Zhang, X. and Zhou, J. (2009) Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field, *FEMS Microbiol. Ecol.*, **70**, 324-333.
- Mason, O.U., Di Meo-Savoie, C.A., Van Nostrand, J.D., Zhou, J., Fisk, M.R. and Giovannoni, S.J. (2008) Prokaryotic diversity, distribution, and insights into their role in biogeochemical cycling in marine basalts, *ISME J.*, **3**, 231-242.
- Milton, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D., Mone, A., Gagne, G., Meier, H., Peyretaillade, E. and Peyret, P. (2007) PhylArray: phylogenetic probe design algorithm for microarray, *Bioinformatics*, **23**, 2550-2557.

- Nordberg, E.K. (2005) YODA: selecting signature oligonucleotides, *Bioinformatics*, **21**, 1365-1370.
- Pozhitkov, A.E., Tautz, D. and Noble, P.A. (2007) Oligonucleotide microarrays: widely applied poorly understood, *Brief. Funct. Genomic. Proteomic.*, **6**, 141-148.
- Rhee, S.-K., Liu, X., Wu, L., Chong, S.C., Wan, X. and Zhou, J. (2004) Detection of Genes Involved in Biodegradation and Biotransformation in Microbial Communities by Using 50-Mer Oligonucleotide Microarrays, *Appl. Environ. Microbiol.*, **70**, 4303-4317.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, **22**, 4673-4680.
- Van Nostrand, J.D., Wu, W.-M., Wu, L., Deng, Y., Carley, J., Carroll, S., He, Z., Gu, B., Luo, J., Criddle, C.S., Watson, D.B., Jardine, P.M., Marsh, T.L., Tiedje, J.M., Hazen, T.C. and Zhou, J. (2009) GeoChip-based analysis of functional microbial communities during the reoxidation of a bio-reduced uranium-contaminated aquifer, *Environ. Microbiol.*, **11**, 2611-2626.
- Vetriani, C., Tran, H.V. and Kerkhof, L.J. (2003) Fingerprinting Microbial Assemblages from the Oxidic/Anoxic Chemocline of the Black Sea, *Appl. Environ. Microbiol.*, **69**, 6481-6488.
- Wang, F., Zhou, H., Meng, J., Peng, X., Jiang, L., Sun, P., Zhang, C., Van Nostrand, J.D., Deng, Y., He, Z., Wu, L., Zhou, J. and Xiao, X. (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent, *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 4840-4845.
- Wang, X. and Seed, B. (2003) Selection of oligonucleotide probes for protein coding sequences, *Bioinformatics*, **19**, 796-802.
- Wernersson, R. and Nielsen, H.B. (2005) OligoWiz 2.0--integrating sequence feature annotation into the design of microarray probes, *Nucleic Acids Res.*, **33**, W611-615.
- Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998) Prokaryotes: The unseen majority, *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 6578-6583.
- Wu, L., Liu, X., Schadt, C.W. and Zhou, J. (2006) Microarray-Based Analysis of Subnanogram Quantities of Microbial Community DNAs by Using Whole-Community Genome Amplification, *Appl. Environ. Microbiol.*, **72**, 4931-4941.
- Yergeau, E., Kang, S., He, Z., Zhou, J. and Kowalchuk, G.A. (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect, *ISME J.*, **1**, 163-179.

SUPPLEMENTARY DATA

A



B

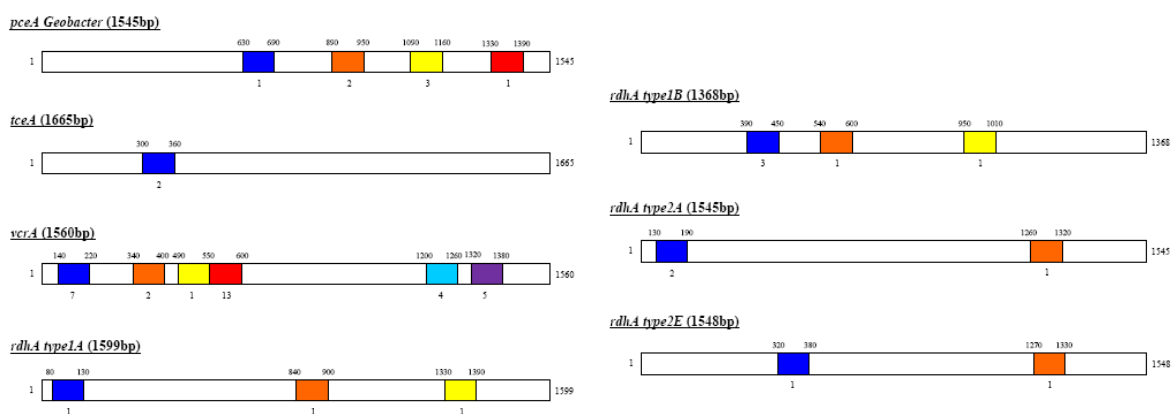


Fig. S1. Gene regions targeted by probes. (A) Individual gene sequences. (B) Consensus degenerate sequences of gene family. The gene length is indicated in brackets. Numbers above genes indicate the location of the gene regions targeted by probes (colored rectangles). Numbers below the gene indicate the number of probes targeting each region.

Table S1. Genes encoding enzymes involved in chlorinated solvent biodegradation targeted by the functional microarray and their accession numbers. Numbers of sequences taken into account for the consensus generation given in brackets.

	Gene name	Organism(s) genus(era)	Accession number(s)
Unique sequence	<i>amoB</i>	<i>Nitrosomonas</i>	L08050 REGION: 1081..2343
	<i>todC1</i>	<i>Pseudomonas</i>	J04996 REGION: 620..1972
	<i>todC2</i>	<i>Pseudomonas</i>	J04996 REGION: 2083..2646
	<i>bmoC</i>	<i>Thauera</i>	AY093933 REGION: 11724..12818
	<i>bvcA</i>	<i>Dehalococcoides</i>	AY563562
	<i>pceA</i> (DET0318)	<i>Dehalococcoides</i>	NC 002936 REGION: 304666..306153
	<i>mmoC</i>	<i>Methylosinus</i>	S81887 REGION: 412..1434
	<i>pceA</i>	<i>Dehalobacter</i>	AJ439607 REGION: 1909..3564
	<i>pceA</i>	<i>Sulfurospirillum</i>	AF022812 REGION: 3272..4777
	<i>pceA</i> (Ssed_2100)	<i>Shewanella</i>	NC 009831 REGION: 2540904..2542328
	<i>prdA</i>	<i>Desulfitobacterium</i>	AB194706 REGION: 141..1532
	<i>rdhA1</i>	<i>Dehalobacter</i>	AJ539532
	<i>rdhA2</i>	<i>Dehalobacter</i>	AJ539533 REGION: 1..1194
	<i>tceA</i>	<i>Dehalococcoides</i>	AF228507 REGION: 825..2489
	<i>vcrA</i>	<i>Dehalococcoides</i>	AY322364 REGION: 623..2182
	<i>xamoA</i>	<i>Xanthobacter</i>	AJ006979 REGION: 38..1531
Consensus sequence	<i>tceA</i> family (22)	<i>Dehalococcoides</i>	AF228507 REGION: 825..2489 ; DQ202132 REGION: 59..1723 ; DQ202131 REGION: 36..1700 ; AY165311 REGION: 47..1711 ; AY165309 REGION: 96..1760 ; AY165312 REGION: 88..1752 ; AY165310 REGION: 103..1767 ; DQ202136 REGION: 98..1762 ; AB274945 REGION: 651..2315 ; AB274951 REGION: 651..2315 ; AB274952 REGION: 651..2315 ; AB274949 REGION: 656..2320 ; AB274955 REGION: 651..2315 ; AB274954 REGION: 656..2320 ; AB274950 REGION: 651..2315 ; AB274948 REGION: 651..2315 ; AB274947 REGION: 651..2315 ; AB274946 REGION: 651..2315 ; DQ202130 REGION: 53..1717 ; DQ202135 REGION: 1..1583 ; DQ202133 REGION: 1..1583 ; DQ202134 REGION: 1..1583
	<i>vcrA</i> family (4)	<i>Dehalococcoides</i>	AY322364 REGION: 623..2182 ; AB268344 REGION: 47..1606 ; EU137842 REGION: 1..1530 ; DQ177519
	<i>pceA</i> family (2)	<i>Geobacter</i>	NC 010814 REGION: 3077378..3078922 ; NC 010814 REGION: 3079858..3081402
	<i>rdhA1A</i> family (3)	<i>Dehalococcoides</i>	NC 002936 REGION: 1067812..1069410 ; NC 009455 REGION: 980352..981950 ; NC 007356 REGION: 885869..887467
	<i>rdhA1B</i> family (3)	<i>Dehalococcoides</i>	NC 002936 REGION: 173382..174749 ; NC 009455 REGION: 173374..174741 ; NC 007356 REGION: 170338..171705
	<i>rdhA2A</i> family (3)	<i>Dehalococcoides</i>	NC 002936 REGION: 290049..291593 ; NC 009455 REGION: 119160..120704 ; NC 007356 REGION: 208160..209704
	<i>rdhA2E</i> family (3)	<i>Dehalococcoides</i>	NC 002936 REGION: 298706..300253 ; NC 009455 REGION: 103582..105129 ; NC 007356 REGION: 72513..74060

Table S2. Primers and PCR amplification conditions. The reaction starts with 5 min of denaturation at 95°C followed by 40 cycles consisting of a denaturation step at 95°C for 30 s, an annealing step for 30 s at the temperature indicated in the table and an elongation step at 72°C for 90 sec.

Gene	Primer pairs	Annealing temperature (°C)	Cycle number	Reference
<i>todC1</i>	36F1_1717-1736: TGGCGCAAGCTGTCTCTGGA 40R1_3140-3159: CTGAAAGCACGTCATCGGCA	64	40	This study
<i>pceA</i>	SH_pceA_1F1_67: GCTGCAGCAACGATAGCTCC SH_pceA_5R1_591: AGCACCTGCCATACGTGCAG	60	40	This study
<i>bvcA</i>	bvcA_73F: GGTGCCGCGACTTCAGTT bvcA_1011R: GTCCTTGATAGCCAAGTGCTTTT	57	40	(Ritalahti <i>et al.</i> , 2006) (Lee <i>et al.</i> , 2008)
<i>vcrA</i>	vcrA_400F: CGGGCGGATGCACTATTTT vcrA_1367R: GGGCAGGAGGATTGACACAT	57	40	(Ritalahti <i>et al.</i> , 2006) (Lee <i>et al.</i> , 2008)

Table S3. Primers and quantitative PCR experiment conditions.

Gene	Primer pair	Annealing temperature (°C)	Reference
<i>pceA</i> (<i>Sulfurospirillum</i>)	pceASulf_154F: GCAATGACTGCAGGTTCTCC pceASulf_223R: ACGCTGTTCGTA CTTCAGCA	59	This study
<i>bvcA</i>	bvcA_517F: TGCCTTGATGGGTCCACTGG vcrA_686R: TCGGTGATTGGCATTACACC	59	This study
<i>vcrA</i>	vcrA_1155F: GATAGCTCCTACCCCGCCAA vcrA_1252R : CTTGGTGCACACCCCGAGAG	59	This study

Table S4. Chlorinated solvent concentrations ($\mu\text{g.l}^{-1}$) measured in the industrial contaminated samples during the biostimulation processes.

Compound	June 2006	June 2007	January 2008	March 2009
Vinyl Chloride	11.3	300	85	280
Cis-1,2-Dichloroethene	1038	4000	1600	3300
Trichloroethene	6725	140	92	12
COHV total	7774.3	4440	1777	3592

Supplementary Data 1. Background noise evaluation method.

Background noise was determined according to “RANDOM probe response” of Nimblegen microarrays. These probes are 20 to 56-mers random oligonucleotide sequences. These are randomly positioned across the microarray surface and can serve as a metric of the background noise (non-specific annealing and background fluorescence). Our method takes

into account the background (B) noise which is characterized by two components: its position (B_{position}) (1) and its dispersion ($B_{\text{dispersion}}$) (2). Moreover, spatial effect across array surface is a predominant within-slide experimental artefact that needs to be eliminated before any other normalization procedure (Wang *et al.*, 2006). That is why for all array images obtained in this work, the surface was segmented to 16 sub-squares according to probe position (X, Y) indicated in pair report. Then, background noise is calculated as follows:

(1) B_{position} is the median intensity of all RANDOM probes considered (in the entire image or in a sub-square).

(2) $B_{\text{dispersion}}$ indicates the variation observed across RANDOM probe intensity values. $B_{\text{dispersion}}$ is more difficult to apprehend than B_{position} since it greatly depends on “image cleanness”.

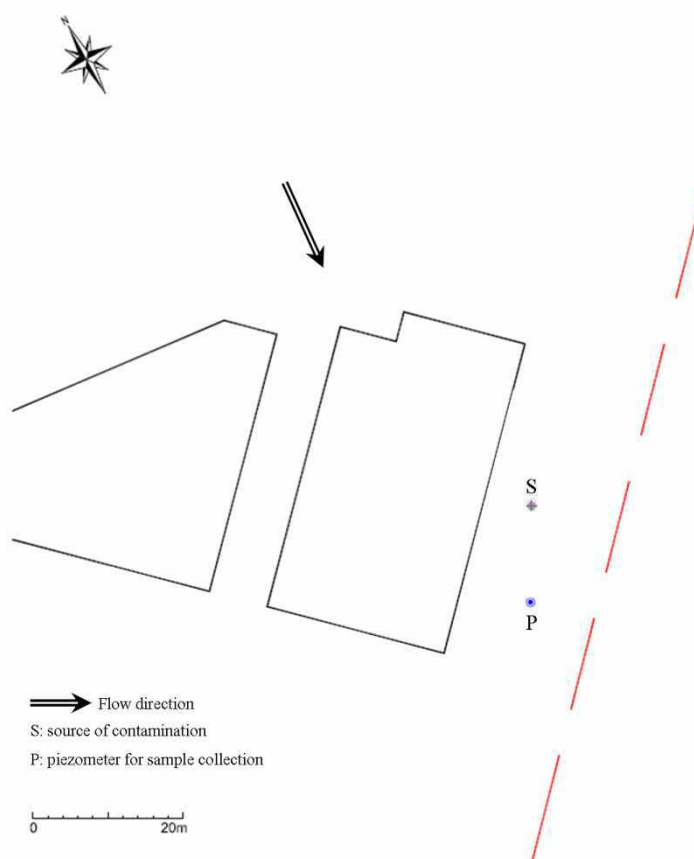
The “image cleanness” is determined and the $B_{\text{dispersion}}$ is calculated according to the variation coefficient (V) observed on all RANDOM probes considered:

- If $V > 33.33\%$: image is considered as “dirty” and ($B_{\text{dispersion}} = B_{\text{position}} - Q3$) where Q3 is the third quartile calculated from all RANDOM probe intensities considered.
- If $V < 33.33\%$: image is considered as “clean” and ($B_{\text{dispersion}} = B_{\text{position}} - D8$) where D8 is the eighth decile calculated from all RANDOM probe intensities considered.

Supplementary Data 2. Description of contaminated site.

The contaminated site has hosted an industrial activity. This has engendered a chlorinated solvent (tetrachloroethene (PCE) and trichloroethene (TCE)) pollution of soils and aquifers. The groundwater, between 7 and 26 m deep, presents an average permeability of $1.7 \times 10^{-4} \text{ m} \cdot \text{sec}^{-1}$, a porosity of 15% and a volume estimated at 9500 m^3 for a surface of 4800 m^2 . TCE is the major contaminant measured in this compartment. The site has been subjected

to three molasse injections into the groundwater since August 2006 (approximately one per year) in order to stimulate anaerobic bacterial populations and activate reductive dechlorination pathways. Water samples were collected during the biostimulation treatment in a piezometer (P) near the source of contamination (S) to measure chlorinated solvent concentrations (Table S2) and at the end of the treatment (March 2009) for molecular experiments.



REFERENCE SUPPLEMENTARY DATA

- Lee, P.K.H., Macbeth, T.W., Sorenson, K.S., Jr., Deeb, R.A. and Alvarez-Cohen, L. (2008) Quantifying genes and transcripts to assess the in situ physiology of "Dehalococcoides" spp. in a trichloroethene-contaminated groundwater site, *Appl. Environ. Microbiol.*, **74**, 2728-2739.
- Ritalahti, K.M., Amos, B.K., Sung, Y., Wu, Q., Koenigsberg, S.S. and Löffler, F.E. (2006) Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains, *Appl. Environ. Microbiol.*, **72**, 2765-2774.
- Wang, X., He, H., Li, L., Chen, R., Deng, X.W. and Li, S. (2006) NMPP: a user-customized NimbleGen microarray data processing pipeline, *Bioinformatics*, **22**, 2955-2957.

4. Discussion

Malgré le nombre important de logiciels de design de sondes pour biopuce ADN (Lemoine *et al.*, 2009), beaucoup ne sont pas adaptés pour répondre à des problématiques environnementales. Le développement de nouveaux logiciels s'avère donc nécessaire pour proposer des stratégies innovantes en relation avec les exigences de ce domaine de recherche. Dans ce contexte, le logiciel HiSpOD offre de nouvelles perspectives en écologie microbienne par comparaison avec les autres outils de détermination de sondes pour FGA, en particulier grâce à la prise en compte d'un plus grand nombre de critères de sélection. Pour exemple, c'est l'un des seuls logiciels, avec Metabolic Design (Terrat *et al.*, 2010 : ARTICLE ANNEXE 1), à pouvoir travailler à partir de séquences dégénérées, paramètre indispensable pour la détermination de sondes exploratoires. Grâce à cette stratégie, il est possible de couvrir non seulement tous les variants de séquences connus pour une famille de gènes donnée, mais aussi de produire des oligonucléotides spécifiques de nouvelles séquences encore jamais observées et dont le nombre sera fonction du niveau de dégénérescence des sondes (comme cela est détaillé dans le chapitre précédent de ce manuscrit). Cependant, l'approche développée au travers d'HiSpOD est différente de celle proposée par Metabolic Design, pour lequel les sondes déterminées après traduction inverse de motifs protéiques sont généralement courtes pour éviter un niveau de dégénérescence trop élevé. Au contraire, HiSpOD n'utilise que des séquences nucléiques en entrée, ce qui le conduit à sélectionner des sondes beaucoup moins dégénérées. Ainsi, bien qu'elles apparaissent moins exploratoires, elles peuvent être plus longues et donc suffisamment sensibles pour la détection de cibles au sein d'échantillons environnementaux. Ce logiciel augmente également les chances de pouvoir identifier plusieurs sondes spécifiques pour chaque gène et donc la confiance dans les résultats obtenus.

La possibilité de tester la spécificité de chaque sonde contre une base de données de CDS spécialisée (plus de 10 millions de séquences), procure au logiciel HiSpOD un avantage supplémentaire non négligeable pour la sélection de sondes très spécifiques. Cependant, ce test de spécificité est l'étape qui impacte le plus sur les temps de calculs nécessaires pour le design. Actuellement installé sur une machine multiprocesseurs de 16 unités centrales de traitement (CPUs), l'évolution prochaine d'HiSpOD concernera la parallélisation à grande échelle de cette étape *via* l'utilisation de ressources informatiques plus importantes comme des clusters ou des grilles de calculs, ce qui devrait conduire à la réduction des temps de calculs et ainsi permettre la détermination de sondes pour un très grand nombre de gènes.

Quoi qu'il en soit, la pertinence de ce nouveau logiciel pour l'élaboration de FGA adaptées à des problématiques environnementales a été validée. Après construction d'une biopuce dédiée à la dégradation des chloroéthènes et son hybridation à l'aide d'un échantillon complexe extrait d'un site contaminé par du TCE, les résultats obtenus ont montré la capacité de cet outil à détecter la présence des gènes impliqués dans toutes les étapes de la déchloration réductrice anaérobie. Ainsi, les conclusions découlant de l'analyse de la FGA confirment les hypothèses élaborées à partir des données chimiques collectées sur le même site qui montraient une déchloration complète des polluants jusqu'en éthylène.

Le logiciel HiSpOD apparaît donc comme un outil performant pour la détermination de sondes présentant tous les critères indispensables pour une application en écologie microbienne : spécificité, sensibilité et caractère exploratoire. La FGA élaborée grâce à cette stratégie présente suffisamment de sensibilité et de spécificité pour être appliquée à l'étude d'environnements complexes. Il est maintenant nécessaire, pour confirmer sa puissance comme outil diagnostic, de l'appliquer non seulement pour l'identification mais aussi le suivi spatio-temporel des populations microbiennes impliquées dans les processus de bioremédiation à partir d'échantillons collectés sur plusieurs sites contaminés.

PARTIE IV :
ETUDE *IN SITU* DES POPULATIONS
DEHALORESPIRANTES IMPLIQUEES DANS LA
DEGRADATION DU TCE SUR DES SITES EN
COURS DE REHABILITATION PAR DES
APPROCHES DE BIOSTIMULATION

1. Contexte

Les solvants halogénés, parmi lesquels figurent les chloroéthènes, sont les principaux polluants industriels retrouvés dans les nappes phréatiques et les eaux souterraines. Rien qu'en France, la base de données BASOL (<http://basol.environnement.gouv.fr/>, Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement) recense environ 750 sites contaminés par ces molécules. Ainsi, en raison des risques potentiels pour l'environnement ou pour la santé humaine, leur décontamination constitue actuellement une préoccupation majeure. Pour cela, les industriels disposent de nombreux procédés de remédiation. Cependant, beaucoup sont encore très coûteux, peu efficaces et difficiles à mettre en œuvre (Travis and Doty, 1990; MacDonald and Kavanaugh, 1994). En effet, ils rencontrent, pour la plupart, des problèmes d'application dus notamment, à l'hétérogénéité des milieux ou à l'inaccessibilité des sources et des polluants (Mackay and Cherry, 1989). Par ailleurs, du fait de leurs caractères agressifs et invasifs pour les écosystèmes traités, ces procédés laissent de plus en plus la place aux techniques de bioremédiation microbienne qui exploitent les capacités naturelles des microorganismes à dégrader les polluants (Vidali, 2001; Megharaj *et al.*, 2011). Parmi les approches existantes, l'ERD (Enhanced Reductive Dechlorination), grâce à l'injection d'une source de carbone et/ou de donneurs d'électrons directement dans la nappe phréatique, vise à biostimuler la microflore déjà présente sur le site contaminé et capable de dégrader les molécules chlorées d'intérêt (He *et al.*, 2002). En instaurant des conditions réductrices et en apportant les éléments indispensables à la croissance des microorganismes, l'ajout de ces substances favorise le processus anaérobie de déchloration réductrice.

La mise en place d'une approche de bioremédiation microbienne efficace nécessite une connaissance approfondie des populations bactériennes et de leurs capacités métaboliques. Ceci est devenu possible grâce à l'essor des approches moléculaires qui ont permis de s'affranchir des techniques culturelles (Lovley, 2003; Xu, 2006; Maphosa *et al.*, 2010) et de développer de nouveaux outils autorisant un accès plus facile à ces données. Ainsi, les biopuces ADN sont parmi les stratégies les plus pertinentes actuellement disponibles, notamment, grâce aux nombreuses évolutions réalisées ces dernières années. En effet, elles permettent désormais non seulement d'identifier les microorganismes et leurs potentiels métaboliques au sein d'un écosystème d'intérêt, mais aussi de suivre facilement leur évolution spatio-temporelle. Par conséquent, dans le cadre de problématiques de bioremédiation, ces approches, utilisées comme outils de diagnostic, s'avèrent

particulièrement bien adaptées pour apprécier la faisabilité d'un procédé de dépollution ainsi que pour évaluer son efficacité. Récemment, une biopuce FGA ciblant l'ensemble des gènes impliqués dans les processus métaboliques de biodégradation des chloroéthènes (Dugat-Bony *et al.*, 2011 : ARTICLE 2) a été élaborée dans notre équipe à l'aide du logiciel HiSpOD. Cet outil, baptisé « DechloArray » comporte désormais 760 sondes ciblant 92 variants de séquences correspondant à 24 familles de gènes. De plus, grâce à un format multiplex, il permet l'analyse simultanée de 24 échantillons.

2. Objectifs

L'objectif de cette étude, soumise dans le journal « Microbial Biotechnology », était de valider l'utilisation de la biopuce FGA « DechloArray » comme outil diagnostic performant et rapide afin de répondre aux besoins des industriels spécialisés dans la bioremédiation, en leur offrant un outil qui leur permette de choisir les stratégies de restauration les mieux adaptées en fonction de la situation de chaque site pollué. Ce travail s'est intégré dans le cadre du programme ANR PRECODD Evasol (Evaluation et Développement du Traitement Anaérobie des Sites Pollués par les Solvants Chlorés, 2008-2010), dont le but était de développer des technologies innovantes, facilement utilisables au niveau industriel, pour évaluer rapidement la faisabilité d'un procédé de bioremédiation. Pour cela, quatre sites industriels français, dont les nappes phréatiques étaient polluées par des chloroéthènes et traités par différents procédés de biostimulation, ont été étudiés. Trois sites, présentant des efficacités de dépollution contrastées (complète ou incomplète), ont été échantillonnés à la fin de leur traitement et analysés pour mieux comprendre l'origine de ces différences. Le quatrième, quant à lui, a fait l'objet d'une étude plus approfondie avec un suivi régulier durant plus d'un an grâce à des prélèvements effectués à partir de plusieurs puits répartis sur l'ensemble du site (zone non polluée, zone source de contamination et panache de diffusion). Dans ce dernier cas, le but était de caractériser la dynamique spatio-temporelle des communautés microbiennes en réponse aux injections de substrats et d'identifier les principaux facteurs gouvernant le développement des espèces clés intervenant dans la biodégradation des polluants. Au total, 30 échantillons provenant de ces différents sites ont été analysés par l'approche FGA.

3. Principaux résultats

L'ensemble des échantillons étudiés au cours de ce travail a fait l'objet d'analyses physico-chimiques et microbiologiques importantes. Grâce à la mesure des paramètres physico-chimiques sur le site suivi au cours du temps, il a été montré que le traitement ERD utilisé, consistant en plusieurs injections de lactate dans la nappe phréatique, avait permis d'instaurer des conditions fortement réductrices, propices à la dégradation anaérobie des chloroéthènes, mais également d'apporter une forte quantité de carbone organique disponible pour les microorganismes. Ceci s'est accompagné d'une diminution importante de la charge en polluants, aussi bien au niveau de la zone source de pollution que du panache de diffusion, avec la détection de l'ensemble des sous-produits de dégradation du TCE (DCE, VC et éthylène). Ces premières données traduisaient l'efficacité du traitement pour stimuler la dégradation complète des chloroéthènes et laissaient supposer l'implication d'un processus biologique. Afin de confirmer ces observations et d'identifier les populations microbiennes impliquées, le matériel génétique extrait des mêmes échantillons a été hybridé sur la biopuce DechloArray. L'analyse des résultats confirme l'origine biologique du processus de dégradation observé et montre l'utilisation d'une seule voie métabolique : la déchloration réductrice anaérobie. Elle met également en évidence une répartition spatiale différente des populations déhalorespirantes qui est fonction à la fois de la charge en TCE ou en sous-produits de dégradation et de la période de prélèvement. Ainsi, pour ce site, les hypothèses découlant de ces observations sont que les espèces affiliées aux genres *Sulfurospirillum* et *Dehalobacter* seraient à l'origine de la dégradation du TCE en DCE dans la zone source de contamination, alors que dans le panache, où les concentrations en TCE sont beaucoup plus faibles, ce serait *Geobacter* qui serait responsable de cette première étape de dégradation. Ceci est déduit de la détection de leur gène *pceA* dans ces différentes zones. Par ailleurs, la détection au niveau du panache des gènes *vcrA* et *bvcA*, codant pour les VC-réductases impliquées dans la dégradation du DCE et du VC en éthylène et trouvés uniquement chez certaines souches appartenant au genre *Dehalococcoides*, confirme les capacités métaboliques de la microflore indigène de ce site à pouvoir réaliser la dégradation complète des chloroéthènes.

Sur les trois autres sites pollués étudiés, deux montraient, au travers des analyses chimiques, une dégradation incomplète des chloroéthènes se traduisant par un blocage de la dégradation au niveau du DCE, alors que le troisième présentait une dégradation complète du TCE jusqu'en éthylène. Ces résultats ont été confortés par les données obtenues avec

l'utilisation de la FGA. En effet, aucun signal d'hybridation n'a pu être détecté pour les gènes impliqués dans la fin de la voie de dégradation à partir des cibles extraites des deux sites bloqués alors que le gène *vcrA* a pu l'être avec les échantillons provenant du site pour lequel la dégradation complète était observée. Ainsi le recoupement de toutes ces données permet de conclure que le blocage au niveau du DCE a pour origine l'absence de certaines souches clés appartenant au genre *Dehalococcoides* et impliquées dans les dernières étapes de déchloration réductrice. Par ailleurs, l'utilisation de la biopuce DechloArray a révélé la présence quasi-systématique de gènes codant pour des RdhA putatives classiquement retrouvées chez des membres du genre *Dehalococcoides* (quel que soit le site mais aussi quel que soit le point de prélèvement) mettant ainsi en évidence l'ubiquité de ce groupe bactérien.

Complete TCE degradation to ethene mediated by a complex dehalorespiring community during an *in situ* biostimulation process

Eric Dugat-Bony,^{1,2} Corinne Biderre-Petit,^{1,2} Faouzi Jaziri,^{4,5} Maude M. David,⁶ Jérémie Denonfoux,^{1,2} Delina Y. Lyon,⁶ Jean-Yves Richard,⁷ Cyrille Curvers,⁷ Delphine Boucher,^{2,3} Timothy M. Vogel,⁶ Eric Peyretailade^{2,3} and Pierre Peyret^{2,3*}

¹Clermont Université, Université Blaise Pascal, Laboratoire Microorganismes: Génome et Environnement, BP 10448, F63000, Clermont-Ferrand ; ²UMR CNRS 6023, Université Blaise Pascal, 63000 Clermont-Ferrand, France ; ³Clermont Université, Université d'Auvergne, Laboratoire Microorganismes: Génome et Environnement, BP 10448, F63000, Clermont-Ferrand ; ⁴Clermont Université, Université Blaise Pascal, LIMOS, BP 10448, F-63000 Clermont-Ferrand ; ⁵UMR CNRS 6158, LIMOS, F-63173 Aubière ; ⁶Environmental Microbial Genomics, Laboratoire Ampère, Ecole Centrale de Lyon, Université de Lyon, 69134 Ecully, France ; ⁷SITA Remediation (SUEZ Environnement company), 11 rue du périgord, 69330 MEYZIEU, France.

* Corresponding author. Mailing address: UMR CNRS/Univ. Blaise Pascal 6023, 63000 Clermont-Ferrand, France. Phone: (33) 473 405139. Fax: (33) 473 407670. E-mail: pierre.peyret@univ-bpclermont.fr.

Running title: *in situ* monitoring of microbial dechlorination

ABSTRACT

In situ enhanced reductive dechlorination (ERD), which is the addition of fermentable organic substrates, is now one of the most often applied approaches to reduce the health and environmental risks associated with the presence of chlorinated solvents in groundwater. However, the success of ERD treatment depends on appropriate subsurface microbial community activity. We used a functional DNA microarray named “DechloArray” to identify the composition of microbial functional gene involved in the dechlorination processes in groundwater samples collected from four contaminated sites during biostimulation treatments. Our results, based on both chemical analyses and the DechloArray, indicate for the first time that complete biodegradation occurs through spatial and temporal variations in dehalorespiring populations belonging to *Sulfurospirillum*, *Dehalobacter*, *Geobacter* and *Dehalococcoides* genera. *Sulfurospirillum* appears the most active in the highly contaminated source zone, while *Geobacter* was only detected in the slightly contaminated downstream zone with *Dehalococcoides* strains harbouring VC-reductase genes. The concomitant detection of both *bvcA* and *vcrA* genes suggests the co-existence of distinct *Dehalococcoides* populations in the contaminated groundwater. The widespread occurrence of *rdhA* transcripts confirms the high diversity for *Dehalococcoides* species. In contrast, on sites where only partial dechlorination with *cis*-DCE accumulation was observed, no key *Dehalococcoides* species were detected.

INTRODUCTION

Chlorinated ethenes, such as perchloroethene (PCE) and trichloroethene (TCE), are among the most frequent contaminants found in groundwater and subsurface ecosystems due to their extensive use as a solvent in industrial activities (metal cleaning and degreasing, dry-cleaning and paint stripping) and as intermediates in the production of other chemicals. PCE and TCE tend to accumulate at the bottom of aquifers where they form dense non-aqueous phase liquids (DNAPLs) because of their high density and their poor solubility in water. These DNAPLs serve as a reservoir for long-term leaching and contamination of the groundwater. Due to their toxicity, PCE and TCE represent a serious risk for human health and the environment (49). Hence, contaminated sites need rehabilitation solutions to completely remove these pollutants. Multiple remediation methods are currently available for treating almost every site whatever its geochemical properties (27, 43). The *ex situ* methods (excavation/pumping) are increasingly being replaced by less invasive *in situ* methods such as biostimulation and bioaugmentation with considerable successes (23, 24, 31, 34, 53).

Among the biostimulation techniques, Enhanced Reductive Dechlorination (ERD) has been successfully applied to many aquifers for the degradation of chlorinated ethenes (18, 50). This approach relies on injection of electron donors or fermentable substrates directly in the groundwater to favour the establishment of reducing conditions required for microbial populations capable of anaerobic reductive dechlorination. Through this approach, the very slow and unstable rates of the natural reductive dechlorination can grow several orders of magnitude. However, the success of ERD treatment depends on the efficiency of each step in the dechlorination process which can be dramatically different depending on the environmental conditions and the microbial populations responsible for the reactions. Indeed, while a wide diversity of bacteria have been shown to catalyze the dechlorination of PCE and TCE to dichloroethene (DCE) (15, 38), only members of *Dehalococcoides* genus are known to be able to reduce DCE and vinyl chloride (VC) to ethene (7, 12, 17, 37). Thus, the reductive dechlorination can stall at an intermediate stage, which typically results in an accumulation of *cis*-DCE and VC daughter products known to be more toxic than PCE and TCE (38). This situation can be relatively transient or virtually permanent if the strains of *Dehalococcoides* and conditions suited to those microorganisms are not present. In the last case, competent microorganisms might be introduced at the site *via* bioaugmentation (43).

Currently, new diagnostic tools are available to help answer the question of whether biostimulation, bioaugmentation or both may be necessary at a chlorinated ethene DNAPL site (35). Indeed, many molecular biological tools were developed to identify

Dehalococcoides species as well as specific reductive dehalogenases (RDases) that have different dechlorinating capabilities. Most commonly, the relationship between successful decontamination and the identification of *Dehalococcoides* has been established by using the gene encoding 16S rRNA (21, 30, 44). However, such analyses are inadequate to assess the potential for dechlorinating activity (54). Such an approach identifies the presence of specific populations, but does not determine whether the detected microbial populations have the specific gene sequences coding for reductive dehalogenases (RDases) responsible of a particular compound biodegradation. New molecular tools designed to detect the well known RDase genes (e.g., *tceA*, *vcra*, and *bvcA*) provide a more complete understanding of dechlorinating capability at a site (39, 45, 55, 56).

The understanding of biodegrading population distribution as well as microbial interactions on contaminated sites is crucial to achieve effective bioremediation strategies. In this study, the full diversity and the spatial distribution of dechlorinating populations were assessed across four French chlorinated solvent-contaminated sites under ERD treatment using a functional microarray (FGA) named “DechloArray” and dedicated to the dechlorination processes as presented in a previous study (10). During this work, this FGA was improved to target all known genes encoding enzymes involved in the three major metabolic biodegradation pathways of chlorinated solvents described in the literature: anaerobic reductive dechlorination (15), aerobic oxidation (36) and aerobic cometabolic oxidation (2). Among the four studied sites, one called B was regularly monitored through approximately one year of biostimulation treatment whereas those called F, G and H were sampled only at the end of their *in situ* biostimulation treatment. In addition to molecular approaches, physico-chemical studies were also conducted on each site to allow a better monitoring of the process. Through the analysis of 30 samples retrieved from these sites showing different pollutant loadings and distinct ERD treatments, this study aimed at i) providing insights into shifts of the spatial and temporal distribution of dechlorinating populations along a contaminated area but also in relation to ERD treatment, and ii) identifying key physico-chemical parameters involved in the changes of bacterial communities. Our results highlight the importance of such survey to better understand the evolution of degrading microorganisms according to their location within a contaminated area and the ERD treatment. This could help to apply more quickly the most effective bioremediation strategy depending on sites characteristics.

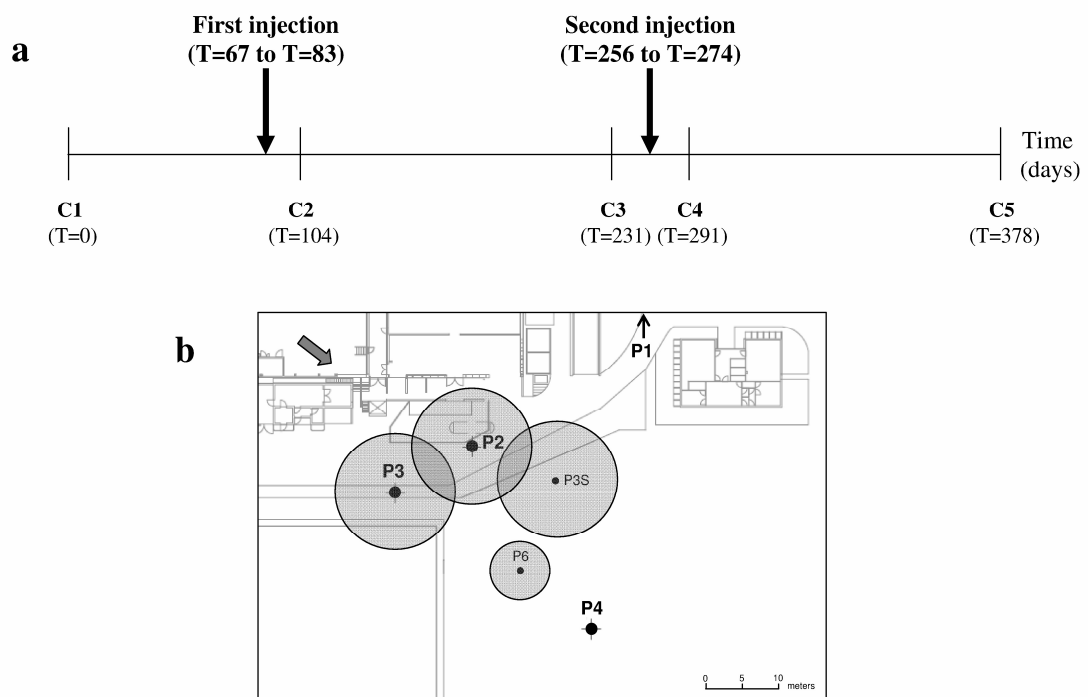


FIG. 1. Overview of ERD strategic scheme applied on site B. (a) Chronological representation of sampling periods and biostimulation treatment by lactate injections. (b) Location of injection and monitoring wells and boundary distribution of lactate immediately after injection (grey circles). P2, P3, P3S and P6: injection wells. P1, P2, P3 and P4: monitoring wells. The upstream monitoring well P1 is located out of the map. The arrow indicated the groundwater flow direction.

MATERIALS AND METHODS

Industrial site description. Four French industrial sites (called B, F, G and H) contaminated with chlorinated solvents were investigated (industrial activities and specific geographical locations not available due to confidentiality agreement).

(i) Site B and ERD demonstration description. Site B was mainly contaminated with an unknown quantity of TCE, resulting from repeated accidental overflowing of a solvent storage tank between 2003 and 2007. TCE levels detected at the source in the underlying groundwater were closed to one hundred mg.l^{-1} (solubility of TCE in water = 1 g.l^{-1}), suggesting the formation of DNAPLs. The subsurface lithology was composed of successive layers of sandy and clay loams. The groundwater, located at 3 m under the surface, was >40 m thick. As the permeability was estimated at $8 \times 10^{-6} \text{ m.s}^{-1}$ and the mean porosity on the site was 2%, the deduced groundwater velocity was calculated at 0.35 m.d^{-1} .

ERD, consisting of two injections of lactate at 6 month interval (at days 67 and 256 respectively, each time over a period of 16-18 days), was applied to target the deep aquifer (Fig. 1a). Based on the volume to be treated (nearly 6000 m^3) and the porosity, 360 kg of lactate in solution (final volume of 140 m^3) were directly injected in four injection wells (30% of volume in P2, 30% in P3, 30% in P3S and 10% in P6) drilled around the contamination source (Fig. 1b). Theoretically, this covered the entire source zone with approximately 3 g.L^{-1} of lactate. The estimated impact radius of the injection varied from 5 to 7.5 m depending on the well. Monitoring was performed between May 2009 and June 2010 through a total of 5 sampling periods (Fig. 1a): C1 (T=0), C2 (T=104 days), C3 (T=231 days), C4 (T=291 days) and C5 (T=378 days) at 4 monitoring wells, *i.e.* one located upstream to the contamination source (P1), one in the contamination source (P2) and two downstream to the contamination source (P3 and P4) (Fig. 1b).

(ii) Treatment of sites F, G and H. Sites F and G showed high levels of TCE ($>15 \text{ mg.l}^{-1}$) in the underlying groundwater at the pollution source, the origin of which is unknown. On these sites, biodegradation was enhanced by three injections of molasses in the plume at site F and in the source zone at site G between August 2006 and October 2008. Groundwater samples used for chemical and molecular analysis were collected in July 2009 (almost 10 months after the end of the treatment) from four monitoring wells for site F (P1, P2, P3 and P4) and three for site G (P1, P3 and P4) located as described for site B. Although the physico-chemical assays performed at the end of the bioremediation process confirmed the presence of favourable reductive conditions (Table S1 in the supplemental material), the biostimulation led to a complete biodegradation of pollutants at site F but to an accumulation of *cis*-DCE at

site G (data not shown). In contrast, site H was a former industrial facility where historical practices resulted in contamination with BTEX (Benzene, Toluene, Ethyl-benzene and Xylene), many hydrocarbons and chlorinated solvents. This last site displayed the lowest chloroethene concentrations. Two remediation treatments were applied, the first consisting of a vacuum extraction between 2000 and 2002 (physical treatment), then followed by an injection of an oil/lactate mixture in 2009 (ERD treatment). These treatments resulted in a build-up of *cis*-DCE, but with a lower extend than for site G. Samples were collected in January 2010, almost a year after the ERD treatment, in four monitoring wells (P1, P2, P3 and P4) as described for the other sites.

Groundwater sample collection and analytical methods. At each site, groundwater samples were collected by SITA Remediation Company (Lyon, France) from the monitoring wells drilled in the deep aquifer *i.e.* between -20 and -30 m depth for the site B, -7 and -12 m for the site F, -7 and -10 m for the site G and -3 and -6 m for the site H. Wells were sampled by low-flow purging methods using a peristaltic pump. Three litres of water were collected per well and then stored on ice until processing. Finally, a total of 30 groundwater samples was collected from all sites (19 from site B, 4 from sites F and H respectively, and 3 from site G) for further analysis.

Chlorinated compounds were analyzed by gas chromatography/mass spectrometry (GC/MS Thermo TRACE DSQ, Thermo Fisher Scientific, Villebon sur Yvette, France) through a headspace sampler (150°C for the injection) and a GC column (TRACE TR-V1, 20 m x 0.18 mm x 1µm, Thermo Fisher Scientific, Villebon sur Yvette, France). The column was held for 6 min at 40°C, followed by a ramp rate of 18 °C.min⁻¹ until 100 °C and 40 °C.min⁻¹ until 200 °C. The carrier gas (helium) had a flow rate of 1 ml.min⁻¹.

Gas composition (methane and ethene) was analyzed by gas chromatography with a 7890A Gas Chromatograph (Agilent, Santa Clara, California, USA) equipped with a headspace sampler, a flame ionization detector (FID) and a bonded polystyrene-divinylbenzene (DVB) based column (HP-PLOT Q, 30 m x 0.53mm x 40 µm, Agilent, Santa Clara, California, USA). The following temperature program was used: 5 min at 60°C with the injector at 250 °C. The carrier gas (nitrogen) had a flow rate of 3 ml.min⁻¹.

Total organic carbon (Hendrickson *et al.*), and ions (dissolved bromide, chloride, fluoride, nitrate, nitrite, orthophosphate and sulfate) were analyzed by ALcontrol Laboratories Company (Hawarden, UK) according to the Nederlands Normalisatie-instituut (NEN, www.nen.nl) norms NEN-EN 1484, NEN-EN-ISO 10304 and NEN-EN-ISO 11732.

Temperature, pH, conductivity, dissolved oxygen (DO) and the oxidation-reduction potential (ORP) were measured onsite in groundwater with a portable multiparameter WTW and appropriate probes (VWR International LLC, West Chester, Pennsylvania, USA).

Nucleic acid extraction. Genomic DNA and total RNA were extracted from all groundwater samples according to the experimental protocol described by Dugat-Bony *et al.* (10). Briefly, 5 to 250 ml of water samples (depending on solids concentration) were filtered under vacuum onto a 47 mm 0.2 μ m pore size GTTP filter (Millipore, Billerica, Massachusetts, USA) and subjected to bead beating disruption followed by phenol chloroform extraction. Both genomic DNA (gDNA) and total RNA are simultaneously extracted. To isolate the RNA fraction, the DNA was digested with 4U of DNase I (DNA-free, Ambion, Austin, Texas, USA) at 37°C for 35 min. gDNA and RNA concentrations were measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, North Carolina, USA) and RNA integrity was verified with a Model 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent, Santa Clara, California, USA).

Sample labelling. The BioPrime Total Genomic Labelling System (Invitrogen, Carlsbad, California, USA) was used to label the gDNA with Alexa Fluor 3 or 5 fluorescent dyes. The reactions used 1 μ g of gDNA template following the manufacturer's instructions and multiplied the amount of DNA by at least 6-fold. A separate method was used for RNA samples. First, the mRNAs were enriched by removing 16S and 23S rRNA from 4 μ g of total RNAs using the MICROBExpress™ Bacterial mRNA Enrichment Kit (Ambion, Austin, Texas, USA), a widely used subtractive hybridization method allowing improvement in mRNA detection sensitivity with a good conservation of mRNA relative abundance (19). In a second step, antisense amino-allyl dUTP marked RNAs (Maillard *et al.*) were produced by amplification using the MessageAmp II-Bacteria kit (Ambion, Austin, Texas, USA) and labelled with Cy3 fluorescent dye (GE Healthcare, Chalfont St Giles, UK) following the manufacturer's instructions. aRNA integrity was monitored with the Model 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent, Santa Clara, California, USA).

For both DNA and RNA samples, the effectiveness of the labelling method was assessed by calculating the amount of labelled nucleotides using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, North Carolina, USA).

Microarray experimental method. The Functional Gene Array (FGA) used in this study was an improved version of the fully validated DechloArray previously developed by Dugat-Bony *et al.* (10). The new version contained 760 oligonucleotide probes of 50-mer length and targeted 92 genes sequences corresponding to 24 families (Table S2 in the

supplemental material). Five distinct regions per gene were targeted and each oligonucleotide probe was synthesized *in situ* with eight replicates. Microarrays were produced in a 12 plex format (Roche NimbleGen, Madison, Wisconsin, USA) allowing the simultaneous analysis of 24 samples.

Purified labelled products (2 μ g for gDNA and 4 μ g for aRNA) were used for the microarray hybridization process using the standard Nimblegen methods. The targets were hybridized for 72 h at 42°C using a 4-bay hybridization system (Roche NimbleGen, Madison, Wisconsin, USA). Array washes were performed as recommended by NimbleGen and slides were scanned at 2 μ m resolution using the InnoScan[®] 900 and Mapix[®] software (Innopsys, Carbonne, France). Finally, pixel intensities were extracted using NimbleScan[™] v2.5 software (Roche NimbleGen, Madison, Wisconsin, USA) and pair reports containing signal intensity data for every spot on the array linked to its corresponding probe identifier were generated.

Microarray data normalization and analysis. Specific scripts developed in this study with the Delphi language were used to establish a four-step automated method to analyze the pair reports generated by NimbleScan[™] v2.5 software. The first two steps, namely local background evaluation and intra-array normalization by SNR' calculation, were as described by Dugat-Bony *et al.* (10). The third step allowed inter-array normalization of the result distributions using the quantile method originally described by Bolstad *et al.* (5). Finally, the last step of the method consisted of a gene response evaluation. Each probe showing a median SNR' ≥ 1.5 from the eight replicates was considered a positive response. Furthermore, all genes for which less than 4 distinct regions of the 5 available were detected were considered negative. Finally, gene response level was calculated as the mean of the highest probe response in each region. The microarray data discussed in this publication are available at the GEO web site (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number: GSE28609.

Redundancy analysis (RDA). The RDA was performed with XLSTAT software (<http://www.xlstat.com/>) on data collected on site B for all monitoring wells from all sampling periods except for P3 sample at the period C1 for which no microarray data were available. In the present study, gene response levels were selected as one set of variables (species data) and geochemical data were selected as the corresponding set of variables (environmental data). The environmental data set consisted of values or concentrations obtained for the following parameters: ORP, Conductivity, DO, PCE, TCE, *cis*-DCE, VC, Ethene, Methane, TOC,

TABLE 1. Behaviour of key parameters analyzed for wells P2 and P3 on site B, both used as injection and monitoring wells, before and after lactate injections.

Sampling periods (T in days)	ORP (mV)		DO (mg.l ⁻¹)		Nitrate (mg.l ⁻¹)		Sulfate (mg.l ⁻¹)		Methane (μg.l ⁻¹)		Ethene (μg.l ⁻¹)		TOC (mg.l ⁻¹)	
	P2	P3	P2	P3	P2	P3	P2	P3	P2	P3	P2	P3	P2	P3
C1 (T=0)	541	692.2	1.44	8.08	7.6	27	56	56	0.03	0.02	<0.01	<0.01	6	5
Lactate (T=67)														
C2 (T=104)	-79.9	-268.3	0.6	0.6	<0.2	<0.2	7.6	13	<0.01	<0.01	0.03	<0.01	1000	910
C3 (T=231)	-50	-49	0.87	1.92	<0.45	<0.45	9	18	<0.01	0.03	0.54	0.19	380	7.4
Lactate (T=256)														
C4 (T=291)	-95	-114.6	0.77	1.1	<0.75	<0.75	65	<5	NA	NA	150	92	2000	1200
C5 (T=378)	-91	-173	NA	0.44	<0.75	0.9	56	<5	64	29	1100	1500	74	270

ORP: oxidation-reduction potential, DO: dissolved oxygen, TOC: total organic carbon, NA: data not available.

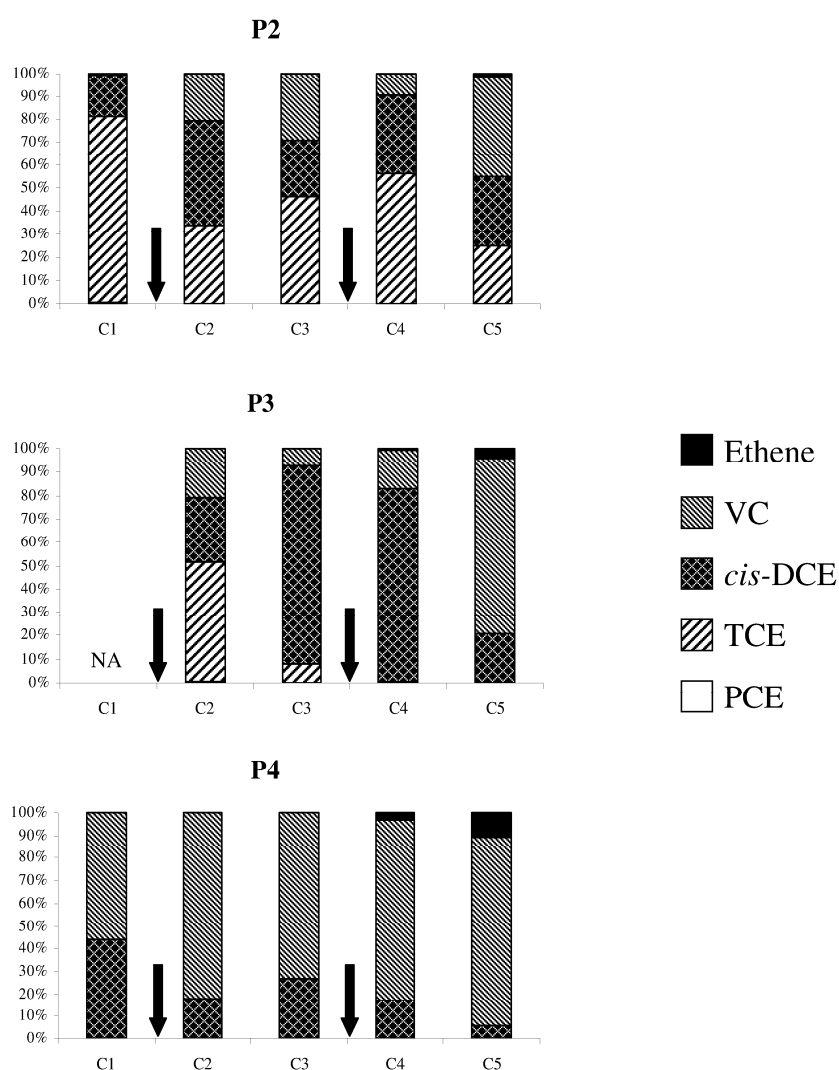


FIG. 2. Chlorinated ethene and ethene trends on site B in the three monitoring wells P2, P3 and P4. Data were provided for sampling periods C1 to C5, when available. Proportion of each compound is expressed in

Nitrate, Sulfate and Chloride. Unrestricted Monte Carlo permutation tests were performed with 499 random permutations and a significance level (P) of 0.05.

RESULTS

Overall impact of the ERD treatment on site B. In order to evaluate the evolution of groundwater geochemistry and electron donors availability during the bioremediation process, analytical measurements were performed on the 19 groundwater samples collected on site B (Table 1). Prior to biostimulation, the groundwater underlying the pollution source area was aerobic with an average redox potential (ORP) of +541 and +692.2 mV in injection wells P2 and P3, respectively. Lactate additions immediately resulted in average ORP values decreasing to -79.9 and -268.3 mV (first injection) and -95 and -114.6 mV (second injection) in both these wells, with dissolved oxygen (DO) consumption indicating the establishment of reducing conditions. With oxygen depletion, we also observed a rapid decrease in nitrate, which could be used first as alternative electron acceptors by certain bacterial populations in response to electron donor loading. In well P3, important decrease in sulfate appeared later than decrease in nitrate and coincided with methane (CH_4) and ethene production. After each injection, lactate was fermented rapidly resulting in an increase of total organic carbon (TOC), which then decreased in the following weeks probably due to the diffusion and consumption of the fermentation products. These results indicated that lactate injections led to the establishment of strongly reductive, oxygen-free and nutrient-rich conditions in the contaminated area.

In parallel, chlorinated ethene and ethene concentration trends were analyzed from the same samples. As expected, only a negligible amount of TCE was detected at the monitoring well P1 located upstream of the source (<1 to $2.5 \mu\text{g.l}^{-1}$ throughout the five sampling periods, data not shown). Similarly, PCE was negligible in all monitoring wells (Fig. 2). Prior to biostimulation, the TCE concentration was greatest at the source well P2 (66 mg.l^{-1}) whereas it was only negligible at the more distant well P4 ($59 \mu\text{g.l}^{-1}$). Upon lactate addition, TCE dechlorination to *cis*-DCE was evident in both P2 and P3 wells. Throughout the experiment, significant reduction in TCE occurred at the source P2 with the average molar fraction decreasing from 80.9% in C1 to 24.7% in C5 whereas it disappeared completely in the nearby well P3 (by sampling period C3) (Fig. 2). In the more distant well P4, the daughter product *cis*-DCE was converted into VC all along the treatment (Fig. 2). In monitoring wells P2, P3 and P4, ethene appeared at the end of the treatment with a higher concentration in well P4 during sampling period C4; this appearance correlated with the onset of CH_4 detection (Fig. 2,

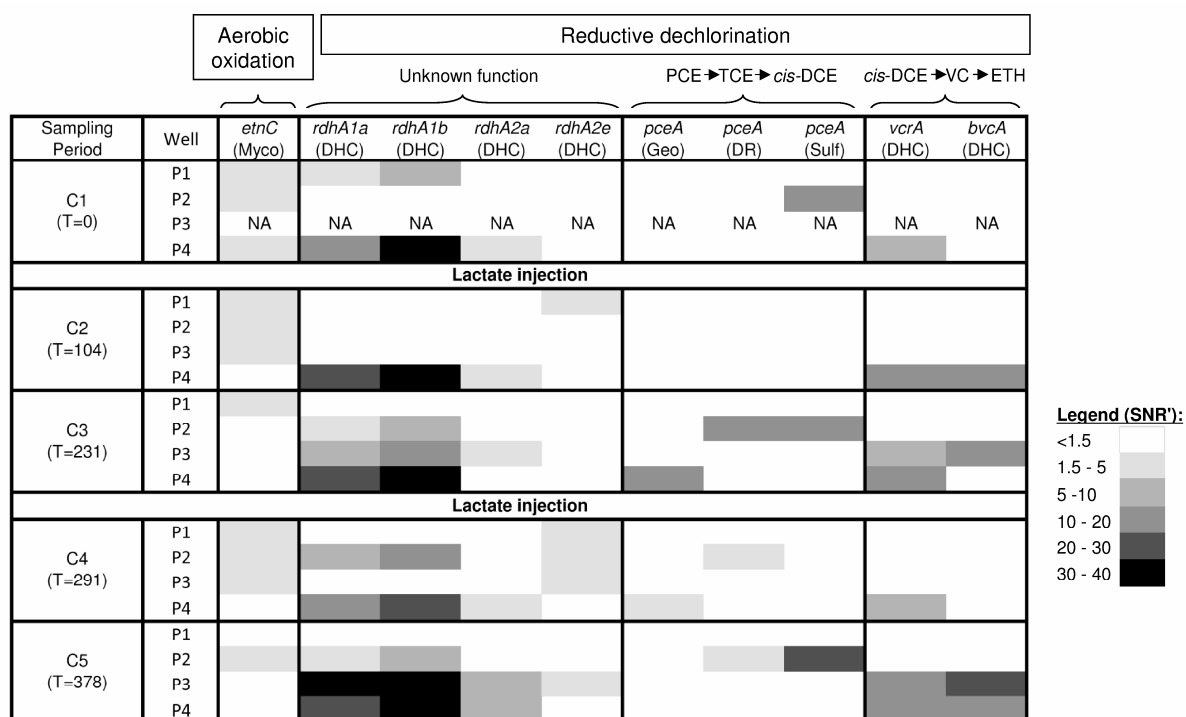


FIG. 3. Gene responses from DechloArray data obtained with gDNA samples collected on site B during ERD demonstration. Gene response levels (mean signal to noise ratio SNR') are represented by coloured cases. Metabolic pathways, gene name and affiliations are indicated for each gene. Myco: *Mycobacterium* ; DHC: *Dehalococcoides* ; Geo: *Geobacter* ; DR: *Dehalobacter* ; Sulf: *Sulfurospirillum*, VC: vinyl chloride, *cis*-DCE: *cis*-dichloroethene, TCE: trichloroethene, PCE: tetrachloroethene, ETH: ethene.

Table 1). Thus, anaerobic conditions created by substrate injections were favourable to chlorinated solvents reductive dechlorination and stimulated the biodegradation in the entire contaminated zone. Furthermore, the rapid evolution of the chloroethene composition clearly reflected the spatial configuration of the contaminated site with (i) a source zone close to the well P2 where the main reaction consists of TCE reduction in its daughter products (*cis*-DCE and VC), and (ii) a plume zone comprising wells P3 and P4 where the essential of the biodegradation consisted in the conversion of *cis*-DCE in VC and ethene, soluble TCE being rapidly consumed.

Molecular detection of functional gene diversity involved in chlorinated ethene degradation on site B. To understand whether the biostimulation process affected the composition and structure of the microbial community involved in dechlorination, the 19 gDNA samples collected on site B during ERD demonstration were hybridized on the DechloArray. The results showed that all genes showing a positive hybridization signal were involved in the reductive dechlorination pathway except the *etnC* gene which encodes an alkene monooxygenase enzyme (Fig. 3). Indeed, this enzyme, isolated from *Mycobacterium* species, is involved in the aerobic oxidation pathway (36) (Table S2 in the supplemental material). This gene was detected without major signal variation even after the lactate injections, in almost all samples including those from the uncontaminated well P1 (Fig. 3). Hence, these findings suggested that *Mycobacterium* species could be present in the indigenous bacterial community but without real involvement in dechlorination success.

At sampling period C1, prior to the biostimulation treatment, several genes were detected (Fig. 3). It allowed to distinguish two dehalorespiring groups with different biodegradation capabilities in distinct zones (Fig. 3): (i) *Sulfurospirillum* (*pceA* gene) in the source well P2, probably responsible of TCE natural attenuation and (ii) *Dehalococcoides* (*vcrA* and other putative *rdhA* genes) in the plume well P4, the only group known to dechlorinate daughter products such as *cis*-DCE or VC, two compounds highly present in this area before the first lactate injection (Fig. 2). The presence of all these genes persisted during the bioremediation process. In addition, *pceA* and *vcrA* genes showed an important increase of signal intensity after lactate injections. These data could highlight the positive impact of the biostimulation treatment on these dechlorinators. Furthermore, as shown in Fig. 3, each lactate addition decreased the number of genes detected during the following sampling periods (C2 and C4) in the injection wells P2 and P3 but not that of genes detected in the distant well P4, used for monitoring and not for injection (Fig. 1b). This led us to conclude

Metabolic Pathway	Gene	Organism genus	P1		P2		P3	
			DNA	RNA	DNA	RNA	DNA	RNA
Aerobic oxidation	<i>etnC</i>	<i>Mycobacterium</i>						
Reductive dechlorination	<i>rdhA1a</i>	<i>Dehalococcoides</i>						
	<i>rdhA1b</i>	<i>Dehalococcoides</i>						
	<i>rdhA2a</i>	<i>Dehalococcoides</i>						
	<i>rdhA2e</i>	<i>Dehalococcoides</i>						
	<i>pceA (DR)</i>	<i>Dehalobacter</i>						
	<i>pceA (Sulf)</i>	<i>Sulfurospirillum</i>						
	<i>vcrA</i>	<i>Dehalococcoides</i>						
	<i>bvcA</i>	<i>Dehalococcoides</i>						

SNR' values:  <1.5  1.5 – 5  5 – 10  10 – 20  20 – 30  30 – 40

FIG. 4. Comparison of signal intensity levels obtained for degradation genes using DNA and RNA targets collected on site B during the sampling period C5. Gene response levels (mean signal to noise ratio SNR') are represented by coloured 760 cases. Metabolic pathway, gene name and affiliation are indicated for each gene.

that the addition of a large amount of substrate resulted in a strong disturbance of the system and the local native microflora for a relatively long period (> 3 weeks).

According to the sample, up to six different RDase genes were simultaneously detected (Fig. 3). Results showed great variations in spatial and temporal distribution of the key RDase genes of known function (*pceA*, *vcrA* and *bvcA*). As the DechloArray was designed to discriminate between bacterial groups based on their functional gene sequences, our data highlighted associations between different dehalorespiring groups, according to the sample location (i.e. *Dehalobacter/Sulfurospirillum* in the source and *Dehalococcoides/Geobacter*, in the plume) and according to the sampling date (i.e. detection of *Dehalobacter* and *Geobacter* only after lactate injections). Thus, these populations could represent the main dechlorinators involved in the decontamination process. In addition, the detection of additional genes encoding putative RDases specific to *Dehalococcoides* species (*rdhA*) in almost all samples could reflect the widespread occurrence and the importance of this functional group in the dechlorination process (Fig. 3).

In order to confirm dechlorinators' role in the decontamination, total RNA was also extracted from groundwater samples. Only three samples (P1, P2 and P3) collected during the last sampling period (C5) gave RNAs in sufficient quantity and quality to be hybridized on the DechloArray. As shown in Fig. 4, the genes detected using the RNA targets were the same than those detected with the gDNA targets, with always a widespread occurrence of the putative *rdhA* genes. In the source zone (P2), the highest signal intensity was obtained for *Sulfurospirillum pceA* gene, so the TCE dechlorination to *cis*-DCE could be mediated by this population, whereas *Dehalobacter* would play only a minor role in this step. At the same location, a slight expression of *vcrA* gene was also observed, which is specific to *Dehalococcoides* species known to transform *cis*-DCE into VC and ethene (Müller et al., 2004). This finding could therefore suggest that at the end of the treatment, *Dehalococcoides* could be associated to *Sulfurospirillum* to perform the complete dechlorination pathway in the source zone. In the well P3, the stronger signal intensity for both *vcrA* and *bvcA* genes suggested that these genes were up-regulated in this zone. These findings were consistent with physico-chemical analysis which showed an earlier onset of ethene at this location (Fig. 2) and highlighted a higher activity of *Dehalococcoides* species carrying these genes. In contrast, the lower signal intensity obtained for the putative *rdhA* genes, also carried by *Dehalococcoides* species, could imply that either they were carried by additional *Dehalococcoides* strains or they were on the same genome of at least one of the *Dehalococcoides* species but differently regulated at the transcriptional level.

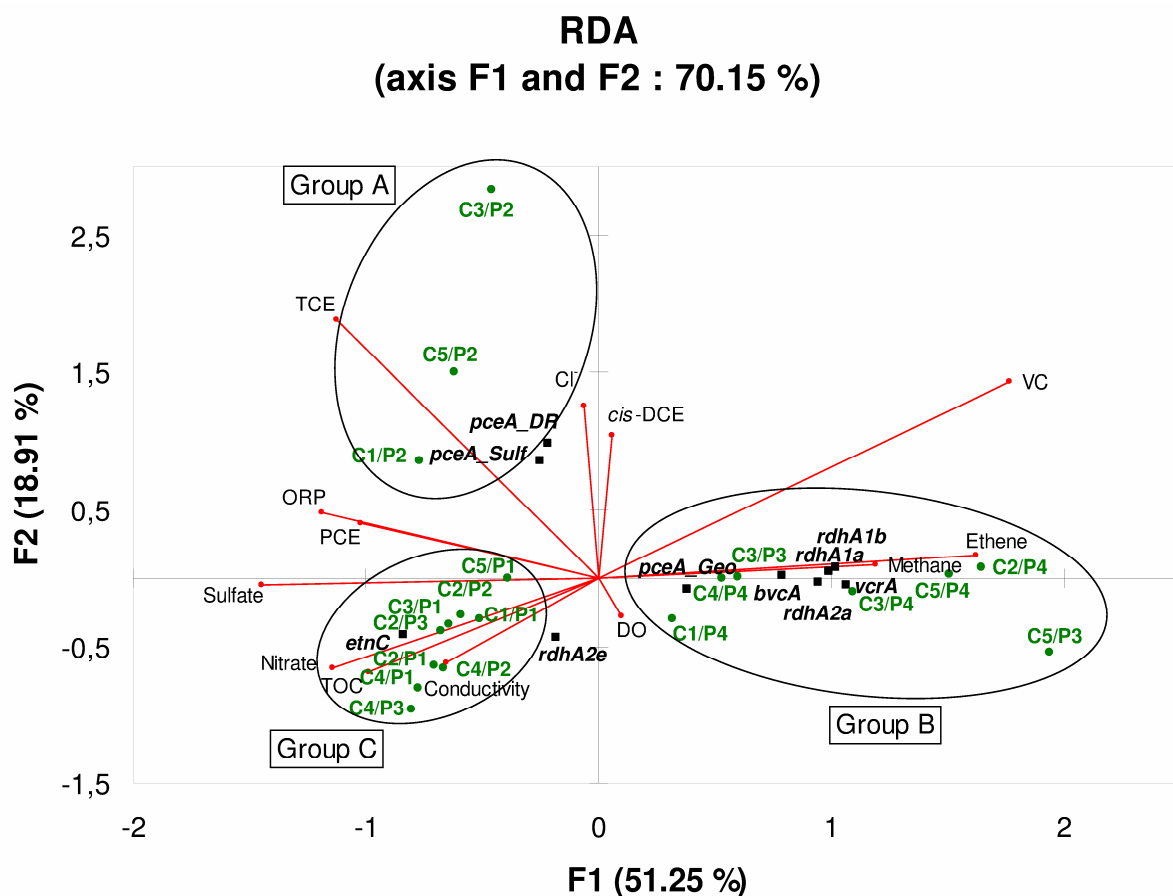


FIG. 5. Redundancy analysis (RDA) of gene response levels as one set of variables (species data) and geochemical data as the explaining set of variables (environmental data) for site B. The length of the arrow is correlated with the strength of relation between the response variables. The arrows point in the direction of the maximum change for the associated variable. Black squares: individual genes. Green stars: groundwater samples. ORP, oxidation reduction potential; TOC, total organic carbon; DO, dissolved oxygen, VC: vinyl chloride, *cis*-DCE: cis-dichloroethene, TCE: trichloroethene, PCE: tetrachloroethene.

Redundancy analysis of gene patterns and geochemical parameters observed on site B. The redundancy analysis (RDA) showed that variation in the detection of genes involved in dechlorination could be explained by different geochemical parameters ($P \geq 0.05$). The first principal canonical axis (F1) accounted for 51.25% of the variation; together with the second canonical axis (F2), this value increased to 70.15% (Fig. 5). On the basis of the sample distribution on the F1 and F2 axis, three sectors were identified. In the first sector (Group A), high TCE concentrations were positively correlated with *Sulfurospirillum* and *Dehalobacter pceA* genes for all samples collected at the source zone P2 that were not perturbed by lactate injections (sampling periods C1, C3 and C5). These genes encode RDases involved in the reduction of both PCE and TCE into *cis*-DCE (33, 40). In contrast, in the second sector (Group B), the three samples unaffected by lactate injections from the plume well P3 (sampling periods C1, C3 and C5), and those from the plume well P4 were significantly associated with high VC and ethene concentrations, and showed strong positive correlation with the detection of VC-reductase genes *vcrA* and *bvcA*. The genes encoding putative RDases (*rdhA1a*, *rdhA1b* and *rdhA2a*) with orthologues in several *Dehalococcoides* genomes (see Table S2 in the supplemental material) were also associated with this group. This second group also showed a strong positive correlation with CH₄ concentration and a strong negative correlation with ORP, sulfate and nitrate concentrations. The sporadic detection of *pceA* from *Geobacter* species only in well P4 could suggest its importance in the removal of TCE traces remaining in this plume. The last sector (Group C) grouped all samples for which no or few degradation potentialities were detected: P1 samples and those from P2 and P3 affected by lactate injections (sampling periods C2 and C4). This last group was also positively correlated with high sulfate and nitrate concentrations. Thus, this analysis revealed that environmental parameters, particularly electron acceptors (nitrate, sulphate, TCE, *cis*-DCE or VC) and reductive conditions (ORP, methanogenesis), controlled the spatial and temporal establishment of key microbial populations involved in the biodegradation process during the bioremediation treatment. In addition, functional gene patterns identified in each sample appeared to be closely related with the degradation reactions occurring in the nearby area, suggesting that these fingerprints could serve as indicator of site conditions and allow optimizing microbial activity for improved bioremediation processes.

Diagnostic of biodegradation capacities on sites F, G and H. Eleven gDNA samples retrieved from different zones on the three additional sites were hybridized on the DechloArray. The results showed that no RDase genes were detected in the upstream well P1 regardless of the site, whereas the *etnC* gene was widely distributed in many wells. In

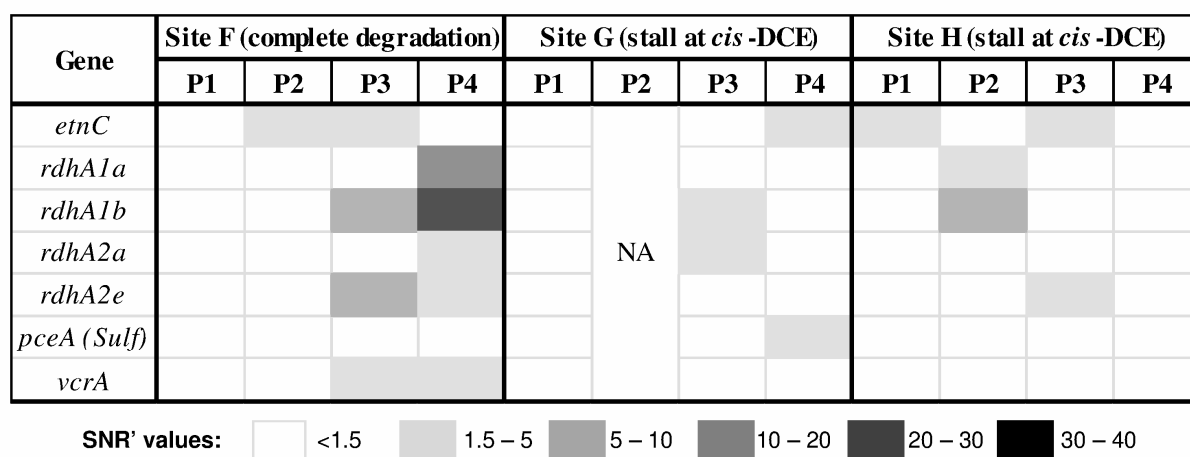


FIG. 6. Gene detection levels observed on sites F, G and H using the DechloArray. Gene response levels (mean signal to noise ratio SNR') are represented by coloured cases. NA: not available.

contrast, the six RDase genes involved in reductive dechlorination already identified from site B, were differently detected in contamination sources and downstream wells, depending on the site (Fig. 6). For site F, both putative *rdhA* and *vcrA* genes were identified in the treated area (wells P3 and P4), reflecting complete dechlorination whereas they were not detected in the non-treated source zone P2 where reducing conditions were not established (negative ORP and slight DO) and the contaminant levels were high (Table S1 in the supplemental material). Surprisingly no known gene able to catalyze TCE dechlorination into *cis*-DCE was detected in wells P3 and P4, which could suggest that this step was realized either by *Dehalococcoides* species using one of their multiple RdhA or by other dechlorinators but not enough abundant to be detectable in this area at the sampling date. For both sites G and H, the apparent stalling of reductive dechlorination at the *cis*-DCE intermediate could be explained by the absence in the plume of all genes involved in the end of the dechlorination process like those encoding VC-reductases (*vcrA* and *bvcA*). Indeed, only the *pceA* gene from *Sulfurospirillum* was detected on site G whereas the putative *rdhA* from *Dehalococcoides* were detected on both sites G and H. In summary, our findings showed that although *Dehalococcoides* populations were widely widespread on the three polluted sites, only site F hosted populations carrying VC-reductase genes that enable dechlorination past *cis*-DCE.

DISCUSSION

ERD treatment efficacy on site B. In this study, ERD was used to cleanup a DNAPL source zone at an industrial contaminated site (Site B). Lactate injections were performed in order to increase the concentration of organic carbon available in the groundwater and to establish reductive conditions favourable to the growth of dechlorinating populations. After almost 1 year of treatment, TCE concentration decreased considerably in the source zone (P2) to be replaced by *cis*-DCE and VC by-products, whereas it disappeared from the plume (wells P3 and P4) concurrent with the appearance of ethene (Fig. 2). However, despite the efficiency of the ERD treatment, the full decontamination of the source zone should take still a quite long time given the TCE quantity discharged and would need additional fermentable substrate injections to maintain favourable conditions for bacterial growth. The application of such long processes with multiple substrate amendments is consistent with what is described by previous studies (24, 28, 34, 52).

Spatial distribution of dechlorinating populations before ERD treatment. RDase functional genes were considered major indicators of dechlorination potential and were recommended for monitoring the presence of known dechlorinators (26, 46). In the present

study, RDase gene diversity was observed on the site B prior to implementation of ERD treatment (Fig. 3), indicating the presence of indigenous dechlorinating consortia. In addition, these dechlorinating populations were heterogeneously distributed with *Dehalococcoides* subgroups identified at several points including both uncontaminated and slightly contaminated zones (P1 and P4), while *Sulfurospirillum* group was only detected in the most contaminated source zone (P2). This distribution suggests a broader natural occurrence of *Dehalococcoides*, which is consistent with previous studies (16, 56). While single *Dehalococcoides* bacterial strains have the capability to completely dechlorinate TCE (37), ethene formation in PCE or TCE-contaminated groundwaters results in most cases from multiple degraders catalyzing the different dechlorination steps. The most frequently found associations involved either *Desulfitobacterium* and *Dehalococcoides* (6, 48, 58), *Dehalobacter* and *Dehalococcoides* (8), *Geobacter* and *Dehalococcoides* (1, 11) or different strains of *Dehalococcoides* (13, 22). In contrast *Sulfurospirillum* and *Dehalococcoides* association was only recently found in an enrichment culture from a bioreactor treating PCE-contaminated groundwater (32) and *in situ* on a French chlorinated solvent contaminated site (10) and seems therefore to be a rare event.

Temporal succession of bacterial communities during ERD treatment. During ERD, it is expected that substrate fermentation creates more favourable environmental conditions to induce growth of the most promising dechlorinating populations. This is supported by our DechloArray data obtained from the *in situ* kinetic experiment performed on site B (Fig. 3). Indeed, soon after the first lactate injection, higher diversity and abundance of RDase genes and microbial populations were observed, with appearance of new dechlorinators belonging to the *Dehalobacter* and *Geobacter* genera. *Dehalobacter*, with *Sulfurospirillum*, was rather located in the highly contaminated source zone, while *Geobacter* was only detected in the slightly contaminated downstream zone (P4) with *Dehalococcoides* strains harbouring VC-reductase genes. As supported by previous studied (3, 9) and by our results (Fig. 5), a such heterogeneous spatial distribution of the microbial community (including dechlorinators) into a contaminated zone could depend on their own specific requirements such as electron acceptors' nature and concentrations (TCE, cis-DCE, VC and potential alternatives nitrate and sulfate), environmental conditions (redox, methanogenesis) or nutrient availability, but also from the inter species competitions. Concerning dehalorespiring populations, it is the first report that shows the co-existence of *Sulfurospirillum*, *Dehalobacter*, *Geobacter* and *Dehalococcoides* species into the same contaminated zone after an ERD treatment.

Focusing on the *Dehalococcoides* group, our findings demonstrated that lactate injections allowed detection of supplemental RDase genes typically found in this group, *i.e.* *bvcA* and other putative *rdhA* genes (Fig. 3). Furthermore, the concomitant detection of both *bvcA* and *vcrA* suggests the co-existence of distinct *Dehalococcoides* populations at some locations in the contaminated groundwater, since no strain was described to host both genes in its genome. This is consistent with previous studies showing that coordinated function of many *Dehalococcoides* species is essential for a more efficient dechlorination (4, 41).

Impact of substrate injections on biodegradation reactions. As shown previously in this study, lactate injection induces the enhancement of the biological dechlorination processes which then continued throughout the treatment. However, the first reaction observed immediately after each injection during sampling periods C2 and C4 is a significant disruption of the groundwater around the injection wells, resulting in disturbances in molecular analyses with the lack of dechlorinating population detection (Fig. 3). Subsequently, the positive effect of this treatment can be observed only after a lag period of several weeks after the injection (sampling periods C3 and C5). This period could be essential to enable the restoration of groundwater conditions favourable to growth of microorganisms, certainly diluted by the massive addition of lactate solution. This finding suggests that substrate addition upstream the contaminated zone could be preferable, leading to a lower disturbance of the environment where dechlorination needs to occur and an acceleration of the decontamination process. This hypothesis was reinforced by what was observed in the most distant plume well P4, not used for injections, which showed both appearance of *bvcA* and higher signal intensity of *vcrA* immediately after the first lactate injection (C2) in the upstream injection wells (Fig. 3). These results reflected the fast diffusion of fermentation products and quick adaptation of degrading populations to environmental changes. For these reasons, it appears crucial to monitor all the reactions that occur in the contaminated zone at several points (spatial coverage) throughout the treatment (temporal coverage).

Identification of active dehalorespiring populations at the end of the ERD treatment. To assess the microbial populations actively involved in chloroethene reduction, we hybridized enriched mRNA samples collected during the last sampling period on the DechloArray. We found that *pceA* gene from *Sulfurospirillum* was the most transcriptionally active at the source zone, whereas no *pceA* transcript for *Dehalobacter* was detected (Fig. 4). This suggests that environmental conditions established at the source zone would be more favourable to the *Sulfurospirillum* metabolism. In contrast, a slight expression of *vcrA* was detected at the same location, while this gene was not detected from the corresponding gDNA

sample. This could indicate that both VC respiring-*Dehalococcoides* and *Sulfurospirillum* strains would be metabolically active in the DNAPL source zone at the end of the treatment and could cooperate for the complete TCE dechlorination. At the well P4, *cis*-DCE and VC consumption were in line with concomitant expression of VC-reductase genes (*vcrA* and *bvcA*). Finally, the widespread occurrence of *rdhA* transcripts in all monitoring wells is in agreement with the high diversity described for *Dehalococcoides* species and with their largest spectrum for organohalides than other dehalorespiring bacteria (25, 42, 51).

Diagnostic application of FGA for bioremediation assistance. Microbial community structures and functions are affected by many environmental factors. Thus, in several case studies, incomplete dechlorination was reported due to inadequate electron donor distribution or absence (or low abundance) of essential dechlorinators (21). This typically results in a build-up of intermediate degradation products such as *cis*-DCE. Among the three additional contaminated sites assessed in the present study, one exhibited complete dechlorination (site F) while the two others showed a stalling of reductive dechlorination (sites G and H). The DechloArray analysis of these samples confirmed the presence of potentially *cis*-DCE dechlorinating species, with the detection of *vcrA* gene, only on site F where complete biodegradation occurred, and its absence on both stalled sites G and H. In addition, chemical and geochemical data obtained during site characterization provided proofs of favourable reductive conditions at least for sites F and G. Therefore, these results suggest the absence of key *Dehalococcoides* species involved in late-stage dechlorination reactions and provide the first line of evidence to favour bioaugmentation application rather biostimulation at both sites G and H to perform complete dechlorination. Indeed, many studies have demonstrated successful decontamination of polluted sites when adding the missing *Dehalococcoides* species (14, 34).

Microbially mediated reduction of chlorinated solvents is a promising strategy for the remediation of highly contaminated groundwater. A better understanding of microbial community structure and functions in relation to environmental conditions is important for designing a successful bioremediation strategy (29). The use of our DechloArray for the degradation gene detection is an excellent approach to assess or predict the performance of an *in situ* bioremediation process more rapidly and more accurately than most other tools. Indeed, microcosm testing needs long-time to reveal biodegradation capabilities and did not reflect performance of microbial populations in real environmental conditions (57). Furthermore, because FGA enables to simultaneously target many genes (92 for the DechloArray), this technique has the unparallel potential to survey more functions than other

molecular tools such as qPCR assays (20). This tool is also better adapted than metagenomic approach for routine studies requiring rapid analysis of many environmental samples, although the latter gives a deeper view of microbial communities (47).

ACKNOWLEDGEMENTS

This work was supported by the grant ID 2598 from the “Agence De l’Environnement et de la Maîtrise de l’Energie” (ADEME, France) and the grant ANR-07-ECOT-005-05 for the program PRECODO Evasol from “Agence Nationale de la Recherche” (ANR, France).

We are grateful to our undergraduate students Gaëtan Guillaume, Stéphane Freitas and Anne-Sophie Yvroud (IUT Génie Biologique, Clermont-Ferrand). We thank the transcriptomic platform of INRA (Crouël, France) for giving access to the microarray hybridization material.

REFERENCES

1. **Amos, B. K., E. J. Suchomel, K. D. Pennell, and F. E. Löffler.** 2009. Spatial and Temporal Distributions of *Geobacter lovleyi* and *Dehalococcoides* spp. during Bioenhanced PCE-NAPL Dissolution. *Environ. Sci. Technol.* **43**:1977-1985.
2. **Arp, D. J., C. M. Yeager, and M. R. Hyman.** 2001. Molecular and cellular fundamentals of aerobic cometabolism of trichloroethylene. *Biodegradation* **12**:81-103.
3. **Becker, J. G.** 2006. A modeling study and implications of competition between *Dehalococcoides* ethenogenes and other tetrachloroethene-respiring bacteria. *Environ. Sci. Technol.* **40**:4473-80.
4. **Behrens, S., M. F. Azizian, P. J. McMurdie, A. Sabalowsky, M. E. Dolan, L. Semprini, and A. M. Spormann.** 2008. Monitoring Abundance and Expression of "Dehalococcoides" Species Chloroethene-Reductive Dehalogenases in a Tetrachloroethene-Dechlorinating Flow Column. *Appl. Environ. Microbiol.* **74**:5695-5703.
5. **Bolstad, B. M., R. A. Irizarry, M. Astrand, and T. P. Speed.** 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**:185-193.
6. **Bunge, M., J. Kleikemper, C. Miniaci, L. Duc, M. G. Muusse, G. Hause, and J. Zeyer.** 2007. Benzoate-driven dehalogenation of chlorinated ethenes in microbial cultures from a contaminated aquifer. *Appl. Microbiol. Biotechnol.* **76**:1447-56.
7. **Cupples, A. M., A. M. Spormann, and P. L. McCarty.** 2003. Growth of a *Dehalococcoides*-Like Microorganism on Vinyl Chloride and cis-Dichloroethene as Electron Acceptors as Determined by Competitive PCR. *Appl. Environ. Microbiol.* **69**:953-959.
8. **Daprato, R. C., F. E. Löffler, and J. B. Hughes.** 2007. Comparative Analysis of Three Tetrachloroethene to Ethene Halorespiring Consortia Suggests Functional Redundancy. *Environ. Sci. Technol.* **41**:2261-2269.
9. **Dowdleit, K., H. Scholz-Muramatsu, R. Miethling-Graff, L. Vigelahn, M. Freygang, A. B. Dohrmann, and C. C. Tebbe.** 2010. Spatial heterogeneity of dechlorinating bacteria and limiting factors for in situ trichloroethene dechlorination revealed by analyses of sediment cores from a polluted field site. *FEMS Microbiol. Ecol.* **71**:444-59.
10. **Dugat-Bony, E., M. Missaoui, E. Peyretailade, C. Biderre-Petit, O. Bouzid, C. Gouinaud, D. Hill, and P. Peyret.** 2011. HiSpOD: probe design for functional DNA microarrays. *Bioinformatics* **27**:641-648.
11. **Duhamel, M., and E. A. Edwards.** 2007. Growth and Yields of Dechlorinators, Acetogens, and Methanogens during Reductive Dechlorination of Chlorinated Ethenes and Dihaloelimination of 1,2-Dichloroethane. *Environ. Sci. Technol.* **41**:2303-2310.

12. **Duhamel, M., K. Mo, and E. A. Edwards.** 2004. Characterization of a Highly Enriched Dehalococcoides-Containing Culture That Grows on Vinyl Chloride and Trichloroethene. *Appl. Environ. Microbiol.* **70**:5538-5545.
13. **Duhamel, M., S. D. Wehr, L. Yu, H. Rizvi, D. Seepersad, S. Dworatzek, E. E. Cox, and E. A. Edwards.** 2002. Comparison of anaerobic dechlorinating enrichment cultures maintained on tetrachloroethene, trichloroethene, cis-dichloroethene and vinyl chloride. *Water Res.* **36**:4193-202.
14. **Ellis, D. E., E. J. Lutz, J. M. Odom, R. J. Buchanan, C. L. Bartlett, M. D. Lee, M. R. Harkness, and K. A. DeWeerd.** 2000. Bioaugmentation for Accelerated In Situ Anaerobic Bioremediation. *Environ. Sci. Technol.* **34**:2254-2260.
15. **Futagami, T., M. Goto, and K. Furukawa.** 2008. Biochemical and genetic bases of dehalorespiration. *Chem. Rec.* **8**:1-12.
16. **Futagami, T., Y. Morono, T. Terada, A. H. Kaksonen, and F. Inagaki.** 2009. Dehalogenation Activities and Distribution of Reductive Dehalogenase Homologous Genes in Marine Subsurface Sediments. *Appl. Environ. Microbiol.* **75**:6905-6909.
17. **He, J., K. M. Ritalahti, M. R. Aiello, and F. E. Löffler.** 2003. Complete Detoxification of Vinyl Chloride by an Anaerobic Enrichment Culture and Identification of the Reductively Dechlorinating Population as a Dehalococcoides Species. *Appl. Environ. Microbiol.* **69**:996-1003.
18. **He, J., Y. Sung, M. E. Dollhopf, B. Z. Fathepure, J. M. Tiedje, and F. E. Löffler.** 2002. Acetate versus Hydrogen as Direct Electron Donors To Stimulate the Microbial Reductive Dechlorination Process at Chloroethene-Contaminated Sites *Environ. Sci. Technol.* **36**:3945-3952.
19. **He, S., O. Wurtzel, K. Singh, J. L. Froula, S. Yilmaz, S. G. Tringe, Z. Wang, F. Chen, E. A. Lindquist, R. Sorek, and P. Hugenholtz.** 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods.* **7**:807-12.
20. **He, Z. L., J. D. Van Nostrand, L. Y. Wu, and J. Z. Zhou.** 2008. Development and application of functional gene arrays for microbial community analysis. *Trans. Nonferrous. Met. Soc. China.* **18**:1319-1327.
21. **Hendrickson, E. R., J. A. Payne, R. M. Young, M. G. Starr, M. P. Perry, S. Fahnestock, D. E. Ellis, and R. C. Ebersole.** 2002. Molecular Analysis of Dehalococcoides 16S Ribosomal DNA from Chloroethene-Contaminated Sites throughout North America and Europe. *Appl. Environ. Microbiol.* **68**:485-495.
22. **Holmes, V. F., J. He, P. K. Lee, and L. Alvarez-Cohen.** 2006. Discrimination of multiple Dehalococcoides strains in a trichloroethene enrichment by quantification of their reductive dehalogenase genes. *Appl. Environ. Microbiol.* **72**:5877-83.
23. **Kao, C. M., Y. L. Chen, S. C. Chen, T. Y. Yeh, and W. S. Wu.** 2003. Enhanced PCE dechlorination by biobarrier systems under different redox conditions. *Water Res.* **37**:4885-4894.
24. **Kennedy, L. G., J. W. Everett, E. Becvar, and D. DeFeo.** 2006. Field-scale demonstration of induced biogeochemical reductive dechlorination at Dover Air Force Base, Dover, Delaware. *J. Contam. Hydrol.* **88**:119-136.
25. **Kube, M., A. Beck, S. H. Zinder, H. Kuhl, R. Reinhardt, and L. Adrian.** 2005. Genome sequence of the chlorinated compound-respiring bacterium Dehalococcoides species strain CBDB1. *Nat. Biotechnol.* **23**:1269-1273.
26. **Lee, P. K. H., T. W. Macbeth, K. S. Sorenson, Jr., R. A. Deeb, and L. Alvarez-Cohen.** 2008. Quantifying Genes and Transcripts To Assess the In Situ Physiology of "Dehalococcoides" spp. in a Trichloroethene-Contaminated Groundwater Site. *Appl. Environ. Microbiol.* **74**:2728-2739.
27. **Lemming, G., M. Z. Hauschild, J. Chambon, P. J. Binning, C. c. Bulle, M. Margni, and P. L. Bjerg.** 2010. Environmental Impacts of Remediation of a Trichloroethene-Contaminated Site: Life Cycle Assessment of Remediation Alternatives. *Environ. Sci. Technol.* **44**:9163-9169.
28. **Lendvay, J. M., F. E. Löffler, M. Dollhopf, M. R. Aiello, G. Daniels, B. Z. Fathepure, M. Gebhard, R. Heine, R. Helton, J. Shi, R. Krajmalnik-Brown, C. L. Major, M. J. Barcelona, E. Petrovskis, R. Hickey, J. M. Tiedje, and P. Adriaens.** 2003. Bioreactive Barriers: A Comparison of Bioaugmentation and Biostimulation for Chlorinated Solvent Remediation. *Environ. Sci. Technol.* **37**:1422-1431.
29. **Lovley, D. R.** 2003. Cleaning up with genomics: applying molecular biology to bioremediation. *Nat. Rev. Micro.* **1**:35-44.
30. **Lu, X., J. T. Wilson, and D. H. Kampbell.** 2006. Relationship between Dehalococcoides DNA in ground water and rates of reductive dechlorination at field scale. *Water Res.* **40**:3131-3140.
31. **Macbeth, T. W., D. E. Cummings, S. Spring, L. M. Petzke, and K. S. Sorenson, Jr.** 2004. Molecular Characterization of a Dechlorinating Community Resulting from In Situ Biostimulation in a Trichloroethene-Contaminated Deep, Fractured Basalt Aquifer and Comparison to a Derivative Laboratory Culture. *Appl. Environ. Microbiol.* **70**:7329-7341.

32. **Maillard, J., M.-P. Charnay, C. Regeard, E. Rohrbach-Brandt, K. Rouzeau-Szynalski, P. Rossi, and C. Holliger.** 2011. Reductive dechlorination of tetrachloroethene by a stepwise catalysis of different organohalide respiring bacteria and reductive dehalogenases. *Biodegradation*:doi: 10.1007/s10532-011-9454-4.
33. **Maillard, J., W. Schumacher, F. Vazquez, C. Regeard, W. R. Hagen, and C. Holliger.** 2003. Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus*. *Appl. Environ. Microbiol.* **69**:4628-38.
34. **Major, D. W., M. L. McMaster, E. E. Cox, E. A. Edwards, S. M. Dworatzek, E. R. Hendrickson, M. G. Starr, J. A. Payne, and L. W. Buonamici.** 2002. Field Demonstration of Successful Bioaugmentation To Achieve Dechlorination of Tetrachloroethene To Ethene. *Environ. Sci. Technol.* **36**:5106-5116.
35. **Maphosa, F., W. M. de Vos, and H. Smidt.** 2010. Exploiting the ecogenomics toolbox for environmental diagnostics of organohalide-respiring bacteria. *Trends. Biotechnol.* **28**:308-316.
36. **Mattes, T. E., A. K. Alexander, and N. V. Coleman.** 2010. Aerobic biodegradation of the chloroethenes: pathways, enzymes, ecology, and evolution. *FEMS Microbiol. Rev.* **34**:445-75.
37. **Maymó-Gatell, X., Y.-t. Chien, J. M. Gossett, and S. H. Zinder.** 1997. Isolation of a Bacterium That Reductively Dechlorinates Tetrachloroethene to Ethene. *Science* **276**:1568-1571.
38. **McCarty, P. L.** 1997. Breathing with Chlorinated Solvents. *Science* **276**:1521-1522.
39. **Müller, J. A., B. M. Rosner, G. von Abendroth, G. Meshulam-Simon, P. L. McCarty, and A. M. Spormann.** 2004. Molecular Identification of the Catabolic Vinyl Chloride Reductase from *Dehalococcoides* sp. Strain VS and Its Environmental Distribution. *Appl. Environ. Microbiol.* **70**:4880-4888.
40. **Neumann, A., G. Wohlfarth, and G. Diekert.** 1998. Tetrachloroethene dehalogenase from *Dehalospirillum multivorans*: cloning, sequencing of the encoding genes, and expression of the *pceA* gene in *Escherichia coli*. *J. Bacteriol.* **180**:4140-5.
41. **Nishimura, M., M. Ebisawa, S. Sakihara, A. Kobayashi, T. Nakama, M. Okochi, and M. Yohda.** 2008. Detection and identification of *Dehalococcoides* species responsible for in situ dechlorination of trichloroethene to ethene enhanced by hydrogen-releasing compounds. *Biotechnol. Appl. Biochem.* **051**:1-7.
42. **Nonaka, H., G. Keresztes, Y. Shinoda, Y. Ikenaga, M. Abe, K. Naito, K. Inatomi, K. Furukawa, M. Inui, and H. Yukawa.** 2006. Complete Genome Sequence of the Dehalorespiring Bacterium *Desulfitobacterium hafniense* Y51 and Comparison with *Dehalococcoides ethenogenes* 195. *J. Bacteriol.* **188**:2262-2274.
43. **Pant, P., and S. Pant.** 2010. A review: Advances in microbial remediation of trichloroethylene (TCE). *J. Environ. Sci. (China)* **22**:116-126.
44. **Rahm, B., S. Chauhan, V. Holmes, T. Macbeth, K. Jr. Sorenson, and L. Alvarez-Cohen.** 2006. Molecular characterization of microbial populations at two sites with differing reductive dechlorination abilities. *Biodegradation* **17**:523-534.
45. **Regeard, C., J. Maillard, and C. Holliger.** 2004. Development of degenerate and specific PCR primers for the detection and isolation of known and putative chloroethene reductive dehalogenase genes. *J. Microbiol. Methods.* **56**:107-118.
46. **Ritalahti, K. M., B. K. Amos, Y. Sung, Q. Wu, S. S. Koenigsberg, and F. E. Löffler.** 2006. Quantitative PCR Targeting 16S rRNA and Reductive Dehalogenase Genes Simultaneously Monitors Multiple *Dehalococcoides* Strains. *Appl. Environ. Microbiol.* **72**:2765-2774.
47. **Roh, S. W., G. C. J. Abell, K.-H. Kim, Y.-D. Nam, and J.-W. Bae.** 2010. Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends. Biotechnol.* **28**:291-299.
48. **Rouzeau-Szynalski, K., J. Maillard, and C. Holliger.** 2011. Frequent concomitant presence of *Desulfitobacterium* spp. and "*Dehalococcoides*" spp. in chloroethene-dechlorinating microbial communities. *Appl. Microbiol. Biotechnol.* **90**:361-8.
49. **Ruder, A. M.** 2006. Potential Health Effects of Occupational Chlorinated Solvent Exposure. *Ann. N. Y. Acad. Sci.* **1076**:207-227.
50. **Scheutz, C., N. d. Durant, P. Dennis, M. H. Hansen, T. Jørgensen, R. Jakobsen, E. e. Cox, and P. L. Bjerg.** 2008. Concurrent Ethene Generation and Growth of *Dehalococcoides* Containing Vinyl Chloride Reductive Dehalogenase Genes During an Enhanced Reductive Dechlorination Field Demonstration. *Environ. Sci. Technol.* **42**:9302-9309.
51. **Seshadri, R., L. Adrian, D. E. Fouts, J. A. Eisen, A. M. Phillippy, B. A. Methe, N. L. Ward, W. C. Nelson, R. T. Deboy, H. M. Khouri, J. F. Kolonay, R. J. Dodson, S. C. Daugherty, L. M. Brinkac, S. A. Sullivan, R. Madupu, K. E. Nelson, K. H. Kang, M. Impraim, K. Tran, J. M. Robinson, H.**

- A. Forberger, C. M. Fraser, S. H. Zinder, and J. F. Heidelberg.** 2005. Genome Sequence of the PCE-Dechlorinating Bacterium *Dehalococcoides ethenogenes*. *Science* **307**:105-108.
52. **Song, D. L., M. E. Conrad, K. S. Sorenson, and L. Alvarez-Cohen.** 2002. Stable Carbon Isotope Fractionation during Enhanced In Situ Bioremediation of Trichloroethene. *Environ. Sci. Technol.* **36**:2262-2268.
53. **Sorenson, K. S.** 2002. Enhanced Bioremediation for Treatment of Chlorinated Solvent Residual Source Areas, p. 119-131, *Chlorinated Solvent and DNAPL Remediation*, vol. 837. American Chemical Society.
54. **Taş, N., M. H. A. Van Eekert, W. M. De Vos, and H. Smidt.** 2010. The little bacteria that can – diversity, genomics and ecophysiology of ‘*Dehalococcoides*’ spp. in contaminated environments. *Microb. Biotechnol.* **3**:389-402.
55. **Taş, N., M. H. A. van Eekert, G. Schraa, J. Zhou, W. M. de Vos, and H. Smidt.** 2009. Tracking Functional Guilds: “*Dehalococcoides*” spp. in European River Basins Contaminated with Hexachlorobenzene. *Appl. Environ. Microbiol.* **75**:4696-4704.
56. **van der Zaan, B., F. Hannes, N. Hoekstra, H. Rijnaarts, W. M. de Vos, H. Smidt, and J. Gerritse.** 2010. Correlation of *Dehalococcoides* 16S rRNA and Chloroethene-Reductive Dehalogenase Genes with Geochemical Conditions in Chloroethene-Contaminated Groundwater. *Appl. Environ. Microbiol.* **76**:843-850.
57. **Xu, J.** 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol. Ecol.* **15**:1713-31.
58. **Yang, Y., M. Pesaro, W. Sigler, and J. Zeyer.** 2005. Identification of microorganisms involved in reductive dehalogenation of chlorinated ethenes in an anaerobic microbial community. *Water Res.* **39**:3954-66.

SUPPLEMENTARY MATERIAL

TABLE S1. Physico-chemical data obtained for the three contaminated sites F, G and H. NA: data not available

Sample	Site F				Site G				Site H			
	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4
DO (mg.l ⁻¹)	1.09	0.98	0.63	0.28	1.03	0.27	0.6	0.71	NA	NA	1.87	NA
ORP (mV)	7.6	53.2	-396.5	-436.3	59	-395.4	-381.4	-464.4	NA	NA	NA	NA
pH	5.95	6.32	6.35	7.17	5.85	5.42	5.27	6.24	NA	NA	6.68	NA
Conductivity (μS.cm ⁻¹)	379	205	1118	514	353	458	457	489	NA	NA	808	NA
PCE (μg.l ⁻¹)	4600	16	<1	4.6	<1	84	10	<1	<1	<1	<1	<1
TCE (μg.l ⁻¹)	6500	97000	13	47	3.5	1100	48	72	<1	11	<1	<1
cisDCE (μg.l ⁻¹)	15000	17000	1300	760	23	3100	9100	16000	<10	160	<10	<10
CV (μg.l ⁻¹)	6.9	96	100	410	<0.5	180	230	200	0.5	49	1.9	0.98
Ethene (μg.l ⁻¹)	<0.01	<0.01	<0.01	0.1	<0.01	0.12	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Methane (μg.l ⁻¹)	<0.01	0.02	8.2	4.1	<0.01	6	5.1	0.34	0.05	4.6	1.5	<0.01
TOC (mg.l ⁻¹)	<5	<5	5.9	59	5.7	130	<50	26	<5	1600	840	<5
Nitrate (mg.l ⁻¹)	4.4	0.72	<0.2	<0.2	<0.2	<0.2	<0.2	<0.2	59	<0.45	<0.45	<0.45
Sulfate (mg.l ⁻¹)	32	28	<2	33	32	2.1	<2	22	340	<5	<5	190

TABLE S2. Description of the 92 gene sequences targeted by the improved version of the DechloArray

Gene sequence accession number (GenBank)	Gene name	Protein	Organism
L08050 REGION: 1081..2343	<i>amoB</i>	Ammonia monooxygenase	<i>Nitrosomonas europaea</i>
J04996 REGION: 620..1972	<i>todC1</i>	Toluene dioxygenase	<i>Pseudomonas putida</i>
J04996 REGION: 2083..2646	<i>todC2</i>	Toluene dioxygenase	<i>Pseudomonas putida</i>
AY093933 REGION: 11724..12818	<i>bmoC</i>	Butane monooxygenase	<i>Pseudomonas butanovora</i>
S81887 REGION: 412..1434	<i>mmoC</i>	Methane monooxygenase	<i>Methylosinus trichosporium</i>
AJ006979 REGION: 38..1531	<i>xamoA</i>	Alkene monooxygenase	<i>Xanthobacter autotrophicus</i> Py2
HM134912	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium</i> sp. JS624
HM134909	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium tusciae</i> JS617
DQ264703	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
DQ264664	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
DQ264659	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
DQ264657	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
DQ264721	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium flavescens</i> NBB1
DQ264663	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
DQ264658	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
DQ264653	<i>etnC</i>	Putative alkene monooxygenase	uncultured bacterium
HM134915	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium tusciae</i> TM2
HM134913	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium mageritense</i> JS625

DQ264722	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium fluoranthenorans</i> NBB2
HM134908	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium gadium</i> JS616
DQ264723	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium rhodesiae</i> NBB3
FJ602754 REGION: 3211..4713	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium</i> sp. JS623
HM134911	<i>etnC</i>	Putative alkene monooxygenase	<i>Mycobacterium rhodesiae</i> JS622
FJ602759	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
FJ602756	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
FJ602755	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
FJ602757	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
FJ602754	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
AY243041	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
FJ602758	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS623
AY243043	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium mageritense</i> JS625
AY243038	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium moriokaense</i> JS619
AY243042	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium</i> sp. JS624
AY243037	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium tusciae</i> JS617
AY243036	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium gadium</i> JS616
AY243040	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium rhodesiae</i> JS622
AY243035	<i>etnE</i>	Epoxyalkane coenzyme M transferase	<i>Mycobacterium aichiense</i> JS61
GQ847815	<i>etnE</i>	Epoxyalkane coenzyme M transferase	uncultured bacterium
GQ847816	<i>etnE</i>	Epoxyalkane coenzyme M transferase	uncultured bacterium
GQ847814	<i>etnE</i>	Epoxyalkane coenzyme M transferase	uncultured bacterium
GQ847817	<i>etnE</i>	Epoxyalkane coenzyme M transferase	uncultured bacterium
GQ847813	<i>etnE</i>	Epoxyalkane coenzyme M transferase	uncultured bacterium
GQ847818	<i>etnE</i>	Epoxyalkane coenzyme M transferase	uncultured bacterium
AJ539532	<i>rdhA1</i>	Chloroethene reductive dehalogenase	<i>Dehalobacter restrictus</i>
AJ539533 REGION: 1..1194	<i>rdhA2</i>	Chloroethene reductive dehalogenase	<i>Dehalobacter restrictus</i>
NC_002936 REGION: 1067812..1069410	<i>rdhA1a</i> (DET1171)	Putative reductive dehalogenase	<i>Dehalococcoides ethenogenes</i> 195
NC_009455 REGION: 980352..981950	<i>rdhA1a</i> (DehaBAV1_0988)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. BAV1
NC_007356 REGION: 885869..887467	<i>rdhA1a</i> (cbdb_A1092)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. CBDB1
NC_002936 REGION: 173382..174749	<i>rdhA1b</i> (DET0180)	Putative reductive dehalogenase	<i>Dehalococcoides ethenogenes</i> 195
NC_009455 REGION: 173374..174741	<i>rdhA1b</i> (DehaBAV1_0173)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. BAV1
NC_007356 REGION: 170338..171705	<i>rdhA1b</i> (cbdb_A187)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. CBDB1

NC_002936 REGION: 290049..291593	<i>rdhA2a</i> (DET0302)	Putative reductive dehalogenase	<i>Dehalococcoides ethenogenes</i> 195
NC_009455 REGION: 119160..120704	<i>rdhA2a</i> (DehaBAV1_0121)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. BAV1
NC_007356 REGION: 208160..209704	<i>rdhA2a</i> (cbdb_A238)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. CBDB1
NC_002936 REGION: 298706..300253	<i>rdhA2e</i> (DET0311)	Putative reductive dehalogenase	<i>Dehalococcoides ethenogenes</i> 195
NC_009455 REGION: 103582..105129	<i>rdhA2e</i> (DehaBAV1_1043)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. BAV1
NC_007356 REGION: 72513..74060	<i>rdhA2e</i> (cbdb_A88)	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. CBDB1
AB194706 REGION: 141..1532	<i>prdA</i>	Tetrachloroethene reductive dehalogenase	<i>Desulfotobacterium</i> sp. KBC1
NC_009831 REGION: 2540904..2542328	<i>pceA</i> (Ssed_2100)	Tetrachloroethene reductive dehalogenase	<i>Shewanella sediminis</i> HAW-EB3
NC_010814 REGION: 3077378..3078922	<i>pceA</i> (Glov_2870)	Tetrachloroethene reductive dehalogenase	<i>Geobacter lovleyi</i> SZ
NC_010814 REGION: 3079858..3081402	<i>pceA</i> (Glov_2872)	Tetrachloroethene reductive dehalogenase	<i>Geobacter lovleyi</i> SZ
AJ439607 REGION: 1909..3564	<i>pceA</i>	Tetrachloroethene reductive dehalogenase	<i>Dehalobacter restrictus</i>
AF022812 REGION: 3272..4777	<i>pceA</i>	Tetrachloroethene reductive dehalogenase	<i>Sulfurospirillum multivorans</i>
NC_002936 REGION: 304666..306153	<i>pceA</i> (DET0318)	Tetrachloroethene reductive dehalogenase	<i>Dehalococcoides ethenogenes</i> 195
GU120391	<i>mbrA</i>	Putative reductive dehalogenase	<i>Dehalococcoides</i> sp. MB
AF228507 REGION: 825..2489	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides ethenogenes</i> 195
DQ202132 REGION: 59..1723	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. SNR-PCE
DQ202131 REGION: 36..1700	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. SFR-cis-DCE
AY165311 REGION: 47..1711	<i>tceA</i>	Trichloroethene reductive dehalogenase	bacterium PM-VC1
AY165309 REGION: 96..1760	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. FL2
AY165312 REGION: 88..1752	<i>tceA</i>	Trichloroethene reductive dehalogenase	bacterium RC-VC2
AY165310 REGION: 103..1767	<i>tceA</i>	Trichloroethene reductive dehalogenase	bacterium YK-TCE1
DQ202136 REGION: 98..1762	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. BRS-PCE
AB274945 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274951 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274952 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274949 REGION: 656..2320	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274955 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274954 REGION: 656..2320	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274950 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274948 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274947 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
AB274946 REGION: 651..2315	<i>tceA</i>	Trichloroethene reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
DQ202130 REGION: 53..1717	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. FMC-TCE

DQ202135 REGION: 1..1583	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. CH-PCE
DQ202133 REGION: 1..1583	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. Owls-PCE
DQ202134 REGION: 1..1583	<i>tceA</i>	Trichloroethene reductive dehalogenase	<i>Dehalococcoides</i> sp. LH-PCE
AY322364 REGION: 623..2182	<i>vcrA</i>	Vinyl chloride reductive dehalogenase	<i>Dehalococcoides</i> sp. VS
AB268344 REGION: 47..1606	<i>vcrA</i>	Vinyl chloride reductive dehalogenase	uncultured bacterium
EU137842 REGION: 1..1530	<i>vcrA</i>	Vinyl chloride reductive dehalogenase	uncultured <i>Dehalococcoides</i> sp.
DQ177519	<i>vcrA</i>	Vinyl chloride reductive dehalogenase	<i>Dehalococcoides</i> sp. KB1
AY563562	<i>bvcA</i>	Putative vinyl chloride reductive dehalogenase	<i>Dehalococcoides</i> sp. BAV1

4. Discussion

Les limites des techniques de réhabilitation physico-chimiques ont permis l'essor de la bioremédiation microbienne et son emploi pour la restauration des sites pollués par les molécules organohalogénées. Cette approche présente l'intérêt d'être efficace et d'un coût raisonnable. Les méthodes les plus courantes sont celles basées sur le principe de biostimulation. Ainsi pour une meilleure dégradation des chloroéthènes, il est possible de stimuler la croissance des microorganismes déhalorespirants et plus particulièrement celle des souches affiliées au genre *Dehalococcoides*. En effet, malgré leur présence sur la quasi-totalité des sites contaminés, elles affichent généralement une faible activité du fait de l'absence des conditions optimales pour leur fonctionnement. Ces microorganismes sont pourtant indispensables pour la dégradation complète des chloroéthènes en composé inoffensif comme l'éthylène, car ce sont les seuls à pouvoir réaliser les étapes finales de la voie métabolique de déchloration réductrice (Hendrickson *et al.*, 2002). Comme nous avons pu le constater au cours de cette étude, l'application de traitements de biostimulation *in situ* par ERD permet l'instauration de conditions anaérobies optimales pour leur développement, ce qui se traduit par l'augmentation rapide des activités de biodégradation dès les premières injections de substrat.

Cependant, pour sélectionner les stratégies les mieux adaptées, les professionnels de la dépollution ont besoin d'outils diagnostics performants leur permettant de suivre, au sein du site contaminé, la présence des fonctions métaboliques naturelles nécessaires à la dégradation du polluant. Pour répondre à cette problématique, ils s'appuient de plus en plus sur des technologies faisant appel à des outils moléculaires comme, par exemple, les biopuces ADN. En effet, ces outils permettent l'identification rapide de tous les microorganismes épurateurs connus et présents sur le site. A partir de ces données, les industriels sont alors capables d'évaluer la pertinence de leurs systèmes de bio-traitement et, lorsque cela s'avère nécessaire, de les ajuster ou de les optimiser. C'est la situation que nous avons rencontré au cours de notre travail avec, pour deux sites biostimulés, un blocage du processus de déchloration réductrice au niveau du DCE. L'analyse des résultats de la biopuce DechloArray, obtenus à partir des échantillons extraits de ces sites, confirme l'absence, au sein de la communauté microbienne indigène, des souches indispensables à la réalisation de la dégradation complète des molécules d'intérêt. Ainsi, pour ces cas particuliers, une approche de bioaugmentation avec l'apport des microorganismes adaptés serait plus appropriée. Il est donc évident que ces

outils présentent un avantage indéniable dans l'aide à la décision pour les professionnels de la dépollution.

Les biopuces ADN offrent également l'opportunité de pouvoir suivre l'évolution des populations microbiennes épuratrices tout au long d'un traitement de dépollution. En effet, si l'effet positif d'un traitement sur la charge en polluants peut être observé simplement par des analyses chimiques, celles-ci n'apportent aucun élément d'information quant aux microorganismes impliqués dans ce processus. Dans notre étude, la DechloArray a été appliquée pour suivre l'évolution des populations microbiennes épuratrices au sein d'un site contaminé et soumis à un traitement de biostimulation par ERD. L'analyse des résultats a mis en évidence la présence, au sein de la microflore indigène, de plusieurs souches appartenant à des genres différents, certains impliqués dans les premières étapes du processus de déchloration réductrice comme *Sulfurospirillum*, *Dehalobacter* et *Geobacter* et d'autres, dans les dernières étapes, comme *Dehalococcoides*. Nous avons également montré qu'elles présentaient une répartition spatiale particulière et que leur abondance évoluait en fonction de l'avancée du traitement. Bien que l'association de plusieurs populations déhalorespirantes avait déjà été décrite dans la littérature (Holmes *et al.*, 2006; Daprato *et al.*, 2007; Maillard *et al.*, 2011; Rouzeau-Szynalski *et al.*, 2011), des *consortia* impliquant autant de *taxa* différents n'avaient encore jamais été observés. Ces résultats indiquent que notre biopuce peut permettre aux industriels d'augmenter considérablement leur capacité à identifier de nouvelles associations de microorganismes potentiellement très efficaces dans les processus de biodégradation d'intérêt.

Les études statistiques, réalisées grâce à la confrontation des données physico-chimiques et biologiques acquises au cours de notre travail, révèlent l'importance de certains paramètres environnementaux pour le bon déroulement *in situ* du processus de déchloration réductrice. Elles montrent clairement, par exemple, que la proportion des différents accepteurs d'électrons (TCE, DCE, VC ou d'autres accepteurs alternatifs comme le nitrate et le sulfate), le potentiel redox ou la concentration en carbone organique totale influent fortement sur la distribution spatio-temporelle des microorganismes déhalorespirants. De plus, une forte corrélation est observée entre les productions d'éthylène et de méthane, ce qui indique que les métabolismes de déhalorespiration et de méthanogénèse sont étroitement liés. Ces résultats sont en parfait accord avec ceux de van der Zaan et ses collaborateurs, qui montrent des tendances similaires pour plusieurs sites contaminés aux Pays-Bas, ceci, grâce au suivi de quatre gènes par qPCR et de nombreux paramètres géochimiques (van der Zaan *et al.*, 2010). Ainsi, aujourd'hui, il s'avère évident qu'une bonne compréhension du fonctionnement des

écosystèmes passe obligatoirement par le développement d'approches pluridisciplinaires et d'interactions fortes entre les différents domaines scientifiques, mais aussi dans le cadre de problématiques environnementales spécifiques comme la bioremédiation, par la prise en compte des compétences et de l'expérience des industriels.

CONCLUSION ET PERSPECTIVES

1. Conclusion

Les microorganismes disposent d'un immense réservoir génétique renfermant des capacités métaboliques uniques ce qui leur procure la capacité de s'adapter à n'importe quelles conditions environnementales et de se développer même dans les milieux les plus extrêmes (Guerrero and Berlanga, 2006; Pikuta *et al.*, 2007). Les changements environnementaux brutaux, comme des événements de pollution, ont deux conséquences majeures : l'une est de provoquer la disparition des organismes sensibles à la perturbation ; l'autre est de contraindre l'évolution et la diversification des organismes tolérants grâce à l'utilisation ou à l'acquisition de nouvelles fonctions leur permettant, par exemple, d'intégrer les molécules polluantes dans leur métabolisme. En effet, chez les bactéries impliquées dans les processus de biodégradation, comme les *Dehalococcoides* par exemple, les gènes cataboliques sont très souvent sujets à des transferts latéraux et/ou associés à des éléments transposables (Krajmalnik-Brown *et al.*, 2007; McMurdie *et al.*, 2009). L'exploitation de ces nouvelles capacités de biodégradation de molécules chimiques suscite un vif intérêt de la part des professionnels de la dépollution. En effet, depuis maintenant plusieurs années, les exemples de traitements réussis à l'aide d'approches de bioremédiation se multiplient et des microorganismes capables de dégrader la plupart des polluants ont été isolés (Dua *et al.*, 2002; Rieger *et al.*, 2002). Sur certains sites contaminés, la dégradation des polluants se fait par atténuation naturelle grâce aux populations indigènes sans que l'homme n'ait besoin d'intervenir. Cependant, bien souvent, l'activité de ces populations est trop faible et doit être soit stimulée par l'apport de substrats, soit compensée par l'apport de souches microbiennes mieux adaptées (Tyagi *et al.*, 2011). Pour la dépollution des aquifères contaminés par des chloroéthènes, le choix entre ces deux stratégies n'est pas automatique. En effet, la biostimulation d'une microflore, au sein de laquelle les populations bactériennes assurant des étapes clés du processus de biodégradation sont absentes, peut conduire à l'accumulation de sous-produits de dégradation non désirables (Hendrickson *et al.*, 2002). Quant à la bioaugmentation, son succès n'est pas systématique car il est fortement dépendant de l'implantation des souches introduites et de leur sensibilité aux conditions géochimiques du site. Ceci est d'autant plus vrai lorsque les souches utilisées sont anaérobies strictes, comme celles appartenant au genre *Dehalococcoides*, puisque l'anaérobiose doit être maintenue tout au long du procédé d'injection (Löffler and Edwards, 2006). Pour ces raisons, les industriels sont demandeurs d'outils de diagnostic et de suivi performants leur permettant de choisir le

traitement approprié aux conditions de terrain mais aussi de suivre l'évolution de la microflore au cours du temps (Lovley, 2003; Maphosa *et al.*, 2010).

Plusieurs techniques moléculaires permettent la caractérisation des communautés microbiennes au sein même d'un environnement complexe. Après avoir étudié les avantages et les inconvénients de chacune d'entre elles, il nous est apparu évident que l'approche biopuce ADN fonctionnelle (FGA), en permettant une détection rapide et simultanée de tous les gènes impliqués dans la biodégradation des polluants, était la stratégie la mieux adaptée pour une application en bioremédiation. Pour autant, face à l'absence de logiciel de sélection de sondes adapté aux problématiques d'écologie microbienne, le développement de nouveaux outils bioinformatiques était nécessaire. En effet, les sondes déterminées doivent remplir les critères de sensibilité et de spécificité, mais aussi être exploratoires pour cibler aussi bien les séquences connues que celles encore inconnues. HiSpOD a été conçu pour tenir compte de l'ensemble de ces critères grâce à la prise en considération d'un plus grand nombre de paramètres et l'intégration d'une base de données représentative des séquences pouvant être rencontrées dans les différents environnements. La phase de validation expérimentale, avec la construction d'une biopuce FGA, a confirmé la pertinence des sondes sélectionnées à l'aide de cet outil et a montré son caractère semi-quantitatif.

Les travaux menés au cours de cette thèse ont conduit à l'élaboration d'une biopuce ADN centrée sur la détection et le suivi de l'ensemble des gènes connus à ce jour pour être impliqués dans les principales voies de biodégradation des chloroéthènes. Cet outil, appelé DechloArray, présente toutes les caractéristiques nécessaires à son application pour l'étude des capacités métaboliques des microorganismes directement à partir d'échantillons environnementaux. Il répond particulièrement bien aux attentes des acteurs de la dépollution, leur offrant un excellent outil pour le diagnostic et pour l'aide à la décision dans le choix de bioprocédé à mettre en œuvre pour une bioremédiation efficace. La version disponible actuellement comporte 760 sondes ciblant 92 variants de séquences pour 24 familles de gènes. Elle couvre ainsi les voies de co-métabolisme aérobie, d'assimilation aérobie et de déchloration réductrice anaérobie. Elle a été produite grâce à une technologie de synthèse *in situ* haute densité sous un format comportant 12 réseaux indépendants afin de pouvoir analyser simultanément jusqu'à 24 échantillons, deux échantillons marqués avec un fluorochrome différent pouvant être hybridés sur chaque réseau. L'ensemble de ces caractéristiques fait que cet outil peut être qualifié de haut débit. Par comparaison, une approche par qPCR absolue ciblant le même nombre de variants nécessiterait au minimum la

préparation de 92 gammes étalons (en général, 8 échantillons de concentrations connues par gamme) soit 736 réactions pour un seul échantillon. Or, comme l'hybridation d'une seule biopuce autorise l'analyse simultanée de 24 échantillons, il faut compter en qPCR, 2208 réactions supplémentaires pour arriver au même résultat. Ceci, d'une part, est techniquement impossible dans un temps identique à celui demandé pour l'analyse par biopuce ADN, et d'autre part, représente un coût largement supérieur. De plus, un algorithme d'analyse d'image couplé à l'outil DechloArray a été développé pour fournir en quelques minutes les résultats d'intensité de signal obtenu pour chaque gène présent au sein de chaque échantillon. Le traitement des données est donc simple et beaucoup plus rapide que celui nécessaire pour d'autres approches haut débit comme la métagénomique. Notre outil peut donc être utilisé à présent en routine dans le cadre de problématiques liées à la bioremédiation.

Outre l'avancée technologique importante ayant conduit au développement de la biopuce DechloArray, son utilisation apporte également des éléments d'information permettant de bien comprendre l'adaptation des populations bactériennes épuratrices aux cours des procédés de bioremédiation et l'influence de certains facteurs environnementaux sur leur développement. En effet, nous avons pu observer, sur des sites contaminés par des chloroéthènes, la mise en place d'associations entre différents microorganismes déhalorespirants, chacun portant ses propres capacités métaboliques. Ces associations, qui peuvent varier d'un site à l'autre comme le montrent ce travail et les nombreuses études menées sur des échantillons provenant d'autres sites pollués (Holmes *et al.*, 2006; Amos *et al.*, 2009; Maillard *et al.*, 2011; Rouzeau-Szynalski *et al.*, 2011), semblent être guidées par la composition de la microflore indigène mais également par les concentrations en différents donneurs et accepteurs d'électrons, par la (les) source(s) de nutriments et par le potentiel redox. La détection du groupe des *Dehalococcoides*, présent sur l'ensemble des sites étudiés au cours de cette étude, n'est pas une garantie de succès pour l'élimination complète des polluants. En effet, seules quelques souches ont acquis les gènes essentiels pour les dernières étapes de la voie de déchloration réductrice. La détection de nombreux autres gènes codant pour des RdhA dont la fonction reste encore inconnue, traduit cependant l'énorme potentiel de ces microorganismes en vue d'applications nouvelles en bioremédiation. C'est pourquoi, l'évolution de notre outil pourrait permettre d'élargir son champ de détection, par exemple pour l'identification des capacités métaboliques intervenant dans la dégradation d'autres molécules polluantes. De plus, il serait également adapté pour répondre à des problématiques écologiques beaucoup plus générales puisque l'on sait maintenant que la nature produit une

très large gamme de composés halogénés mais que très peu de données sont disponibles sur leur cycle naturel et leur impact sur le bon fonctionnement des écosystèmes naturels.

2. Perspectives

Les résultats obtenus tout au long de cette thèse ouvrent de réelles perspectives de travail dans plusieurs domaines de recherche. Les approches pluridisciplinaires mises en place ont abouti au développement de nouveaux outils bioinformatiques et moléculaires suffisamment flexibles et puissants pour être appliqués à la compréhension de multiples processus intervenant dans n'importe quelle problématique biologique. Pour autant, ces outils sont encore perfectibles. La fin de ce mémoire sera donc consacrée à quelques suggestions concernant les évolutions envisageables pour améliorer leur efficacité. Elle permettra également de montrer leur potentiel pour de nouvelles applications dans le cadre d'études plus fondamentales.

2.1. Amélioration des outils de sélection de sondes

Les stratégies de design de sondes exploratoires, développées au sein de notre équipe et désormais intégrées dans les logiciels PhylArray (Militon *et al.*, 2007), Metabolic Design (Terrat *et al.*, 2010) et HiSpOD (Dugat-Bony *et al.*, 2011), présentent deux contraintes principales qu'il est maintenant possible de contourner grâce aux récentes évolutions des outils informatiques.

Le premier inconvénient concerne le temps de calcul nécessaire au design de sondes qui peut s'avérer relativement important lorsque le nombre de gènes à traiter est élevé. En effet, les logiciels que nous avons développés génèrent un nombre de sondes bien supérieur à celui produit par les approches ordinaires. Or, l'évaluation de la spécificité de chacune d'entre elles nécessite une étape extrêmement consommatrice en temps de calculs. Par conséquent, le déploiement de nos logiciels sur des architectures parallèles pourrait contribuer à réduire fortement ces délais et ainsi faciliter la conception de biopuces ADN ciblant de plus en plus de gènes ou de microorganismes. PhylArray est en cours de déploiement sur un cluster composé de 240 CPUs hébergé à l'Institut Supérieur d'Informatique, de Modélisation et de leurs Applications (ISIMA, Clermont-Ferrand) ainsi que sur la grille de calcul « Enabling Grid for E-sciences in Europe » (EGEE). Cette grille est une infrastructure réunissant 70

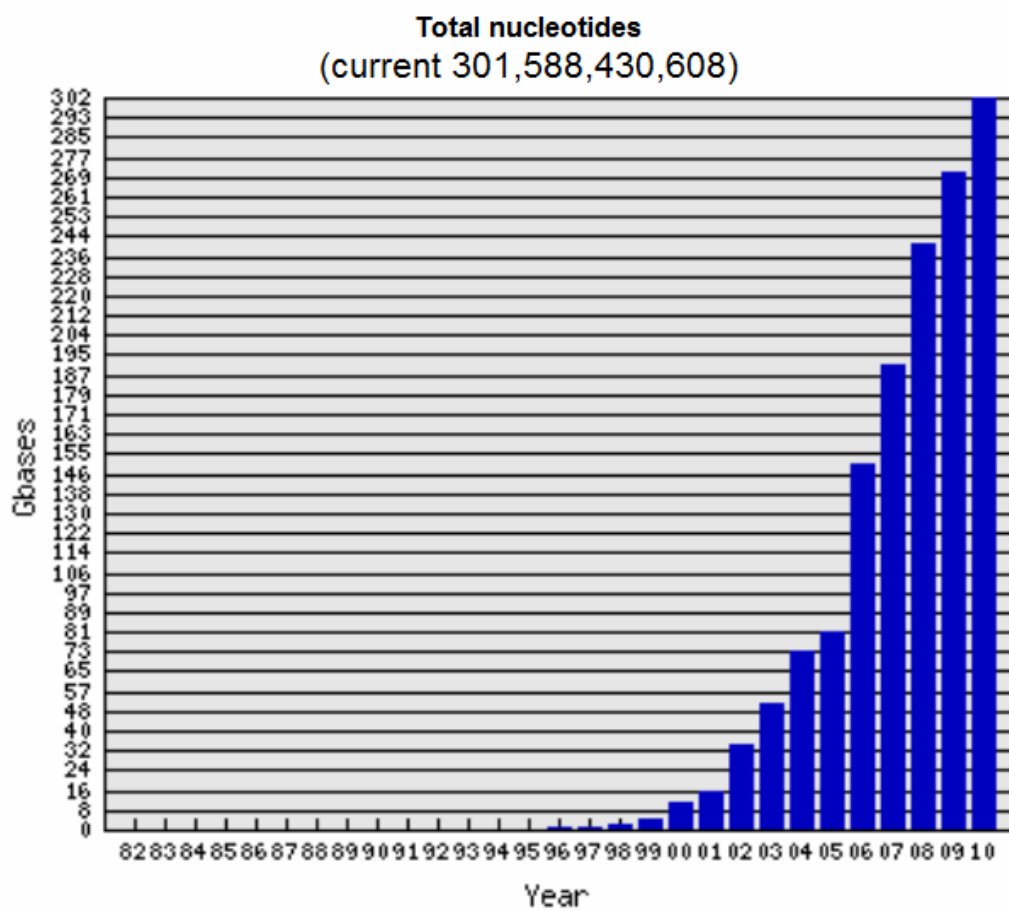


Figure 29 : Représentation graphique de l'évolution du nombre de nucléotides présents dans la base de données EMBL pour la période comprise entre 1982 et 2010 (source : <http://www.ebi.ac.uk/embl/Services/DBStats/>).

organismes partenaires issus des 27 pays de l'Union Européenne et génère une capacité totale de 20 000 CPUs. Le déploiement de Metabolic Design et de HiSpOD sur de telles infrastructures est toujours à l'étude. Cependant, les avancées technologiques sont si rapides que des alternatives intéressantes pour nos applications sont déjà disponibles sur le marché et pourraient constituer des solutions encore mieux adaptées. Par exemple, des calculs parallèles à grande échelle sont réalisables grâce à l'exploitation des processeurs des cartes graphiques (GPUs) de nos propres ordinateurs de travail, bien souvent inutilisés par nos tâches bureautiques quotidiennes, et les premiers outils bioinformatiques basés sur cette technologie commencent à voir le jour (Kohlhoff *et al.*, 2011). Le choix de l'architecture se fera donc en fonction de sa performance et de sa capacité à optimiser l'utilisation de nos logiciels tout en les rendant accessibles au plus grand nombre d'utilisateurs possible.

Par ailleurs, une diminution des temps de calculs pourrait avoir d'autres conséquences positives, offrant notamment la possibilité de considérer plus de critères pour le design. Parmi ces critères, nous pouvons citer les contraintes thermodynamiques et les structures secondaires des sondes et des cibles (Mueckstein *et al.*, 2010). L'influence de ces paramètres, connus pour intervenir sur les réactions d'hybridation au niveau des interfaces liquides / liquides, est encore peu prise en considération pour les expériences de biopuces ADN où les réactions ont lieu à l'interface solide / liquide (Pozhitkov *et al.*, 2007). Ainsi pour atténuer les effets négatifs que pourraient engendrer ces contraintes sur les niveaux de détection perçus, la stratégie de design employée à l'heure actuelle consiste à sélectionner un plus grand nombre de sondes pour cibler différentes régions de chaque gène (Chou *et al.*, 2004). Or, en fonction du gène ciblé, cette démarche peut s'avérer compliquée. La prise en compte de ces nouveaux paramètres par le logiciel au moment du design pourrait donc simplifier l'étape de design en limitant le nombre de sondes nécessaires pour détecter chaque gène.

Le second inconvénient concerne l'absence d'automatisation de la mise à jour des bases de données de qualité utilisées par nos logiciels pour la réalisation des tests de spécificité. Avec l'évolution constante des bases de données internationales comme GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), EMBL (<http://www.ebi.ac.uk/embl/>) et DDBJ (<http://www.ddbj.nig.ac.jp/>) (Figure 29), une mise à jour régulière des bases de données intégrées dans nos logiciels est obligatoire pour que les résultats des tests de spécificité effectués pour chaque sonde soient pertinents. Il serait également judicieux que, parallèlement à une mise à jour de la base de données, soit faite une mise à jour des designs antérieurs afin d'inclure les nouvelles séquences apparues dans les bases de données et d'offrir les biopuces

ADN les plus performantes possibles. Les outils qui seront proposés seront alors entièrement automatisés et actualisés en permanence.

2.2. Généralisation de l'utilisation de nos logiciels

Les logiciels développés dans notre équipe sont modulables et autorisent par conséquent le développement de biopuces ADN adaptées à n'importe quelle problématique biologique.

La FGA DechloArray élaborée au cours de ce travail de thèse est exclusivement centrée sur la thématique « chloroéthènes ». Or, ces composés chimiques ne sont qu'une partie des polluants retrouvés dans l'environnement. En effet, de nombreuses autres molécules organiques et inorganiques contaminent les écosystèmes naturels. A l'image de la biopuce *BiodegPhyloChip*, proposée récemment et ciblant un grand nombre de gènes impliqués dans la biodégradation de divers polluants (Pathak *et al.*, 2011), il serait intéressant d'enrichir notre biopuce avec de nouvelles sondes, toujours déterminées avec HiSpOD pour assurer leur caractère exploratoire, sondes qui seraient ainsi dirigées contre un plus grand nombre de gènes. Ces derniers pourraient intervenir dans la dégradation des polluants majeurs mais également dans d'autres voies métaboliques associées comme la sulfato-réduction et la méthanogenèse. Cette biopuce que l'on pourrait alors qualifier de « Bioremediation » représenterait un outil de choix pour les professionnels de la dépollution. En effet, malgré sa spécialisation, elle permettrait l'étude de n'importe quel site pollué, qu'il soit contaminé par une seule molécule ou par un mélange de produits.

Contrairement à ce que l'on pourrait penser, l'utilisation de nos logiciels n'est pas réservée uniquement au règne des procaryotes ni à des applications en écologie microbienne. Le domaine ciblé dépendra principalement, en réalité, de la base de données utilisée par le logiciel lors du design. Ainsi, grâce à l'intégration de nouvelles bases spécialisées dans HiSpOD, il sera possible de proposer des biopuces dédiées à des eucaryotes comme par exemple *Drosophila melanogaster* (travail en cours) ou encore à des biotopes particuliers comme le sol, l'eau douce, les environnements marins, les systèmes digestifs, etc.... Par ailleurs, une dernière évolution, également extrêmement judicieuse, serait de coupler systématiquement l'analyse d'un échantillon par FGA (fonctionnelle) avec l'hybridation d'une POA (phylogénétique). Ce type d'étude permettrait d'identifier, au sein d'un écosystème d'intérêt, l'ensemble des espèces présentes et l'ensemble de leurs capacités métaboliques en seule expérimentation.

2.3. Etude des processus de déhalorespiration dans les milieux naturels : identification de nouvelles souches et de nouvelles capacités métaboliques

L'intérêt d'étudier les composés chlorés n'entre pas seulement dans une problématique de pollution de l'environnement liée à l'activité humaine. Il existe un grand nombre de raisons d'explorer de façon plus approfondie l'ensemble des processus en relation avec ces molécules. En effet, le chlore, un des éléments chimiques les plus abondants sur notre planète, est essentiel à la vie (Oberg, 2002). Par conséquent, il est retrouvé à l'état naturel dans une grande variété de molécules dont le cycle doit être finement régulé pour maintenir un certain équilibre. Au cours de l'évolution, quelques groupes bactériens se sont donc spécialisés, adaptant leur métabolisme à cette catégorie de molécules (source de carbone, source d'énergie, donneur d'électrons). Ce sont eux qui sont aujourd'hui exploités dans les procédés de bioremédiation. A l'heure actuelle, les microorganismes qui semblent être les plus spécialisés dans l'utilisation de molécules organochlorées sont les bactéries déhalorespirantes appartenant au phylum des *Chloroflexi* (*Dehalococcoides*, *Dehalogenimonas* et DLG). Elles possèdent d'énormes capacités de déchloration qui ont déjà été largement présentées dans ce mémoire. Leur détection quasi systématique dans différents milieux naturels (nappes phréatiques, sédiments de rivières, environnements marins etc...), même si elles ne représentent qu'une fraction minime de la communauté microbienne, indique que ces bactéries occupent des niches écologiques importantes au sein du cycle des halogènes. Or, ce dernier n'est que très rarement pris en compte pour la compréhension du fonctionnement global des écosystèmes.

Comme nous l'avons constaté au cours de ce travail, ces microorganismes suscitent un vif intérêt de la part des industriels et de plus en plus d'informations sont disponibles quant à leur mode de vie dans les milieux anthropisés. En revanche, les données concernant leur écologie dans les milieux non perturbés sont quasi inexistantes, et ceci, malgré leur très large distribution et leur rôle primordial dans l'équilibre de ces milieux. Pour combler ce manque, le projet Microflex financé par le conseil de recherche européen (ERC) a pour principal objectif d'isoler, à partir d'environnements marins, de nouveaux membres appartenant au groupe des *Chloroflexi* et proches des *Dehalococcoides* afin de mieux comprendre leur physiologie (Adrian, 2009). En revanche, aucune recherche ne fait état, à ce jour, d'études comparables au niveau des systèmes d'eau douce.



Figure 30 : Photographie aérienne du lac Pavin (Puy de Dôme, France).

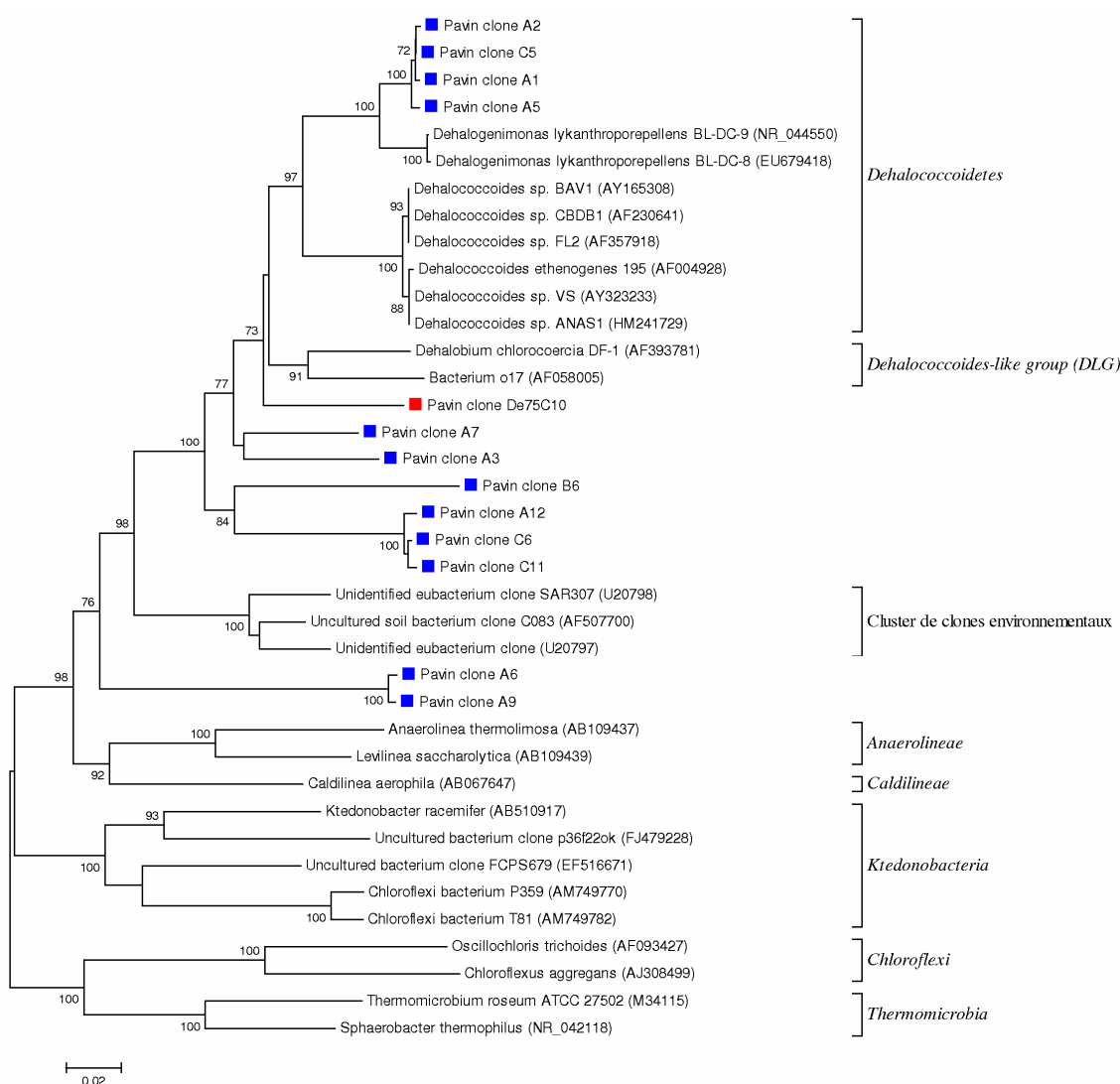


Figure 31 : Arbre phylogénétique représentant les huit groupes du phylum des *Chloroflexi* et intégrant les séquences isolées de la zone anoxique du lac Pavin. Carré rouge : séquence du clone De75C10 obtenu après amplification avec un couple d'amorces universelles ciblant l'ADNr 16S bactérien (Biderre-Petit *et al.*, 2010). Carrés bleus : séquences des clones obtenus avec un couple d'amorces ciblant l'ADNr 16S des *Dehalococcoides*. L'arbre a été construit avec le logiciel MEGA 4 (Tamura *et al.*, 2007) grâce à l'application de la méthode des plus proches voisins (Neighbour Joining) et d'un bootstrap de 100.

2.3.1. Les milieux d'eau douce, sources de nouvelles capacités métaboliques

Le lac Pavin est situé dans une zone de moyenne montagne non urbanisée et non anthropisée du Massif Central (Figure 30). C'est un lac de cratère faisant 92 mètres de profondeur de type méromictique, c'est-à-dire caractérisé par la présence d'une zone anoxique permanente, qui dans ce cas particulier, s'étend sur plus de trente mètres. Des approches moléculaires généralistes entreprises par notre équipe pour étudier les cycles du méthane et du soufre dans cet écosystème, ont révélé la présence de populations minoritaires affiliées au phylum des *Chloroflexi*, dont certaines proches des *Dehalococcoides*, et localisées dans la zone anoxique (Biderre-Petit *et al.*, 2010; Biderre-Petit *et al.*, 2011) (ARTICLE ANNEXE 2).

A partir de ces données, et dans le cadre de cette thèse, une étude plus approfondie utilisant des approches PCR ciblant l'ADNr 16S a été réalisée, ce qui a permis de mettre en évidence une plus grande diversité de séquences formant différents clusters affiliés à la famille des *Dehalococcoidetes* ou au groupe des DLG (Figure 31). L'un de ces clusters fait partie du même embranchement que les deux souches appartenant au genre *Dehalogenimonas* connues pour être capables d'utiliser de nombreux chloroalcanes comme accepteurs d'électrons (Yan *et al.*, 2009a). L'amplification de ce biomarqueur à partir de l'ADNg, mais également à partir de l'ARN, prouve l'activité de ces populations dans le lac Pavin. Enfin, des couples d'amorces supplémentaires ciblant des RdhA ont été utilisés et ont conduit à l'identification de séquences partielles codant pour une protéine proche de VcrA (environ 98% d'identité protéique) qui est impliquée dans les étapes finales de dégradation des chloroéthènes chez plusieurs souches appartenant au genre *Dehalococcoides* (Muller *et al.*, 2004; Sung *et al.*, 2006a; Lee *et al.*, 2011). Tous ces éléments montrent la richesse de cet écosystème en populations potentiellement déhalorespirantes (Figure 31), mais également les capacités de ces dernières à pouvoir dégrader des composés chlorés. Par ailleurs, l'éloignement des séquences identifiées par rapport aux séquences référencées dans les bases de données tend à supposer que les souches habitant ce milieu d'eau douce sont différentes et pourraient donc posséder de nouvelles capacités métaboliques. L'utilisation des outils développés au cours de mon travail de thèse, associés à d'autres approches complémentaires, pourrait permettre d'apporter des éléments d'informations fondamentaux supplémentaires pour comprendre le rôle de ces populations dans le fonctionnement de cet écosystème naturel. Par ailleurs, l'absence de pollution et d'industrie proches de ce milieu lacustre exclut l'origine humaine de substances pouvant être utilisées comme source d'énergie par ces

microorganismes. Cependant, la présence d'un couvert végétal important à proximité du lac (Figure 30) et son origine volcanique sont autant de sources potentielles de molécules organohalogénées dont l'identité reste à déterminer.

2.3.2. Identification de nouvelles potentialités métaboliques : intérêts biotechnologiques

Les outils moléculaires sont indispensables pour l'étude des communautés microbiennes d'écosystèmes complexes. Dans le cadre de ce travail, ils ont permis de montrer la présence de nouvelles souches déhalorespirantes potentiellement intéressantes pour la dégradation de molécules organohalogénées au sein d'un écosystème non perturbé. Cependant, une meilleure connaissance du métabolisme de ces souches passe par la mise en place d'approches complémentaires permettant leur isolement et leur culture. Des essais d'enrichissement de souches ou *consortia* à partir d'échantillons d'eau collectés dans la zone anoxique du lac Pavin et amendés avec des molécules organochlorées sont actuellement en cours au laboratoire. Parallèlement, des analyses chimiques sont envisagées afin d'identifier les molécules naturelles pouvant être utilisées par ces populations et d'orienter les stratégies d'isolement mises en œuvre. Cependant, l'émission de ces molécules dans les environnements naturels conduit généralement à de faibles concentrations locales ce qui complique leur détection et leur identification. L'étude des communautés microbiennes présentes dans ces cultures et de leur dynamique au cours du temps, grâce à l'utilisation d'approches moléculaires comme la qPCR ou le FISH, permettra d'évaluer l'efficacité des techniques culturales. De même, l'étude des signatures isotopiques des molécules chlorées pourrait être mise en place afin de déterminer la nature biotique ou abiotique des processus intervenant dans leur dégradation (Meckenstock *et al.*, 2004), ou de différencier l'action de certaines populations microbiennes (Dong *et al.*, 2011; Fletcher *et al.*, 2011).

Le principal objectif recherché dans l'isolement de ces souches est de pouvoir accéder à l'ensemble de leur matériel génétique. Il sera alors possible d'effectuer une affiliation plus fine des souches isolées mais aussi de caractériser l'ensemble de leurs capacités métaboliques. En effet, ces informations pourraient donner accès à de nouvelles enzymes RdhA possédant des activités intéressantes pour le développement de bioprocédés innovants permettant, par exemple, la dégradation de composés chlorés persistants comme certains pesticides (par exemple le dichlorodiphényltrichloroéthane, DDT ou le chlordécone), les PCDD/Fs et les PCBs). D'un point de vue fondamental, des approches de génomique comparative,

envisageables grâce à la disponibilité de plusieurs génomes de souches déhalorespirantes isolées de milieux contaminés, pourraient permettre de mieux comprendre l'évolution de ces microorganismes mais également les mécanismes d'adaptation qu'ils ont développés pour pouvoir utiliser une gamme aussi large de molécules organohalogénées dans leur processus respiratoire.

Ainsi, les résultats obtenus au cours de cette thèse, avec le développement d'outils innovants et l'acquisition de connaissances nouvelles sur l'organisation des populations microbiennes soit lors de processus de réhabilitation ou soit au sein d'un milieu non perturbé, permettent d'ouvrir des perspectives de recherche excitantes. En effet, elles pourraient conduire à mieux comprendre les mécanismes d'adaptation des microorganismes face à l'introduction, que ce soit lié à l'activité humaine ou non, de molécules halogénées dans leur habitat.

REFERENCES

- Abrahamsson, K, Ekdahl, A, Collen, J et al.** (1995) Marine algae : a source of trichloroethylene and perchloroethylene. *Limnol Oceanogr* **40**: 1321-1326.
- Acierno, D., Barba, A.A., and D'Amore, M.** (2003) *Microwaves in soil remediation from VOCs. 1: Heat and mass transfer aspects*. Hoboken, NJ, ETATS-UNIS: Wiley.
- Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F.** (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**: 8966-8969.
- Adams, J.A., Reddy, K.R., and Tekola, L.** (2011) Remediation of chlorinated solvent plumes using in-situ air sparging-a 2-d laboratory study. *Int J Environ Res Public Health* **8**: 2226-2239.
- Adrian, L.** (2009) ERC-group microflex: microbiology of Dehalococcoides-like Chloroflexi. *Rev Environ Sci Biotechnol* **8**: 225-229.
- Adrian, L., Szewzyk, U., Wecke, J., and Gorisch, H.** (2000) Bacterial dehalorespiration with chlorinated benzenes. *Nature* **408**: 580-583.
- Adrian, L., Hansen, S.K., Fung, J.M., Gorisch, H., and Zinder, S.H.** (2007) Growth of Dehalococcoides strains with chlorophenols as electron acceptors. *Environ Sci Technol* **41**: 2318-2323.
- Ahn, Y.B., Haggblom, M.M., and Kerkhof, L.J.** (2007) Comparison of anaerobic microbial communities from Estuarine sediments amended with halogenated compounds to enhance dechlorination of 1,2,3,4-tetrachlorodibenzo-p-dioxin. *FEMS Microbiol Ecol* **61**: 362-371.
- Ahn, Y.B., Liu, F., Fennell, D.E., and Haggblom, M.M.** (2008) Biostimulation and bioaugmentation to enhance dechlorination of polychlorinated dibenzo-p-dioxins in contaminated sediments. *FEMS Microbiol Ecol* **66**: 271-281.
- Alvarez-Cohen, L., and Speitel, G.E., Jr.** (2001) Kinetics of aerobic cometabolism of chlorinated solvents. *Biodegradation* **12**: 105-126.
- Alvarez-Cohen, L., McCarty, P.L., Boulygina, E., Hanson, R.S., Brusseau, G.A., and Tsien, H.C.** (1992) Characterization of a methane-utilizing bacterium from a bacterial consortium that rapidly degrades trichloroethylene and chloroform. *Appl Environ Microbiol* **58**: 1886-1893.
- Amann, R.I., Krumholz, L., and Stahl, D.A.** (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* **172**: 762-770.
- Amann, R.I., Ludwig, W., and Schleifer, K.H.** (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.
- Amos, B.K., Suchomel, E.J., Pennell, K.D., and Löffler, F.E.** (2009) Spatial and Temporal Distributions of Geobacter lovleyi and Dehalococcoides spp. during Bioenhanced PCE-NAPL Dissolution. *Environ Sci Technol* **43**: 1977-1985.
- Amos, B.K., Ritalahti, K.M., Cruz-Garcia, C., Padilla-Crespo, E., and Löffler, F.E.** (2008) Oxygen effect on Dehalococcoides viability and biomarker quantification. *Environ Sci Technol* **42**: 5718-5726.
- An, Y.J., Kampbell, D.H., Weaver, J.W., Wilson, J.T., and Jeong, S.W.** (2004) Natural attenuation of trichloroethene and its degradation products at a lake-shore site. *Environ Pollut* **130**: 325-335.
- Andersen, M.E., Reddy, M.B., and Plotzke, K.P.** (2008) Are highly lipophilic volatile compounds expected to bioaccumulate with repeated exposures? *Toxicol Lett* **179**: 85-92.
- Ansorge, W.J.** (2009) Next-generation DNA sequencing techniques. *Nat Biotechnol* **25**: 195-203.
- Arciero, D., Vannelli, T., Logan, M., and Hooper, A.B.** (1989) Degradation of trichloroethylene by the ammonia-oxidizing bacterium Nitrosomonas europaea. *Biochem Biophys Res Commun* **159**: 640-643.

- Arp, D.J., Yeager, C.M., and Hyman, M.R.** (2001) Molecular and cellular fundamentals of aerobic cometabolism of trichloroethylene. *Biodegradation* **12**: 81-103.
- Aulenta, F., Canosa, A., Reale, P., Rossetti, S., Panero, S., and Majone, M.** (2009) Microbial reductive dechlorination of trichloroethene to ethene with electrodes serving as electron donors without the external addition of redox mediators. *Biotechnol Bioeng* **103**: 85-91.
- Azizian, M.F., Marshall, I.P., Behrens, S., Spormann, A.M., and Semprini, L.** (2010) Comparison of lactate, formate, and propionate as hydrogen donors for the reductive dehalogenation of trichloroethene in a continuous-flow column. *J Contam Hydrol* **113**: 77-92.
- Barbee, G.C.** (1994) Fate of Chlorinated Aliphatic Hydrocarbons in the Vadose Zone and Ground Water. *Ground Water Monit Remediat* **14**: 129-140.
- Bass, D.H., Hastings, N.A., and Brown, R.A.** (2000) Performance of air sparging systems: a review of case studies. *J Hazard Mater* **72**: 101-119.
- Beeman, R.E., and Bleckmann, C.A.** (2002) Sequential anaerobic-aerobic treatment of an aquifer contaminated by halogenated organics: field results. *J Contam Hydrol* **57**: 147-159.
- Benner, M.L., Mohtar, R.H., and Lee, L.S.** (2002) Factors affecting air sparging remediation systems using field data and numerical simulations. *J Hazard Mater* **95**: 305-329.
- Bertin, P.N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsene-Ploetze, F., Gallien, S. et al.** (2011) Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *Isme J*.
- Biderre-Petit, C., Boucher, D., Kuever, J., Alberic, P., Jezequel, D., Chebance, B. et al.** (2010) Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin) using *aprA* and 16S rRNA gene markers. *Microb Ecol* **61**: 313-327.
- Biderre-Petit, C., Jezequel, D., Dugat-Bony, E., Lopes, F., Kuever, J., Borrel, G. et al.** (2011) Identification of microbial communities involved in the methane cycle of a freshwater meromictic lake. *FEMS Microbiol Ecol* **77**: 533-545.
- Blow, N.** (2008) DNA sequencing: generation next-next. *Nat Meth* **5**: 267-274.
- Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A.** (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5**: 566-582.
- Bonaventura, C., and Johnson, F.M.** (1997) Healthy environments for healthy people: bioremediation today and tomorrow. *Environ Health Perspect* **105 Suppl 1**: 5-20.
- Bontemps, C., Golfier, G., Gris-Liebe, C., Carrere, S., Talini, L., and Boivin-Masson, C.** (2005) Microarray-based detection and typing of the *Rhizobium* nodulation gene *nodC*: potential of DNA arrays to diagnose biological functions of interest. *Appl Environ Microbiol* **71**: 8042-8048.
- Bottari, B., Ercolini, D., Gatti, M., and Neviani, E.** (2006) Application of FISH technology for microbiological analysis: current state and prospects. *Appl Microbiol Biotechnol* **73**: 485-494.
- Bradley, P.M., and Chapelle, F.H.** (1996) Anaerobic Mineralization of Vinyl Chloride in Fe(III)-Reducing, Aquifer Sediments. *Environ Sci Technol* **30**: 2084-2086.
- Bradley, P.M., and Chapelle, F.H.** (1998a) Methane As a Product of Chloroethene Biodegradation under Methanogenic Conditions. *Environ Sci Technol* **33**: 653-656.
- Bradley, P.M., and Chapelle, F.H.** (1998b) Microbial mineralization of VC and DCE under different terminal electron accepting conditions. *Anaerobe* **4**: 81-87.
- Bradley, P.M., and Chapelle, F.H.** (1999) Role for Acetotrophic Methanogens in Methanogenic Biodegradation of Vinyl Chloride. *Environ Sci Technol* **33**: 3473-3476.
- Bradley, P.M., and Chapelle, F.H.** (2000) Acetogenic Microbial Degradation of Vinyl Chloride. *Environ Sci Technol* **34**: 2761-2763.

- Bradley, P.M., Chapelle, F.H., and Lovley, D.R.** (1998a) Humic Acids as Electron Acceptors for Anaerobic Microbial Oxidation of Vinyl Chloride and Dichloroethene. *Appl Environ Microbiol* **64**: 3102-3105.
- Bradley, P.M., Chapelle, F.H., and Lovley, D.R.** (1998b) Humic Acids as Electron Acceptors for Anaerobic Microbial Oxidation of Vinyl Chloride and Dichloroethene. *Applied and Environmental Microbiology* **64**: 3102-3105.
- Bradley, P.M., Landmeyer, J.E., and Dinicola, R.S.** (1998c) Anaerobic Oxidation of [1,2-¹⁴C]Dichloroethene under Mn(IV)-Reducing Conditions. *Appl Environ Microbiol* **64**: 1560-1562.
- Bradley, P.M., Richmond, S., and Chapelle, F.H.** (2005) Chloroethene biodegradation in sediments at 4 degrees C. *Appl Environ Microbiol* **71**: 6414-6417.
- Bravo-Linares, C.M., Mudge, S.M., and Loyola-Sepulveda, R.H.** (2010) Production of volatile organic compounds (VOCs) by temperate macroalgae. The use of solid phase microextraction (SPME) coupled to GC-MS as a method of analysis. *J Chil Chem Soc* **55**: 227-232.
- Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M., and Andersen, G.L.** (2007) Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci U S A* **104**: 299-304.
- Brodie, E.L., Desantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L. et al.** (2006) Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72**: 6288-6298.
- Broholm, K., Jensen, B.K., Christensen, T.H., and Olsen, L.** (1990) Toxicity of 1,1,1-Trichloroethane and Trichloroethene on a Mixed Culture of Methane-Oxidizing Bacteria. *Appl Environ Microbiol* **56**: 2488-2493.
- Bunge, M., and Lechner, U.** (2009) Anaerobic reductive dehalogenation of polychlorinated dioxins. *Appl Microbiol Biotechnol* **84**: 429-444.
- Bunge, M., Wagner, A., Fischer, M., Andreesen, J.R., and Lechner, U.** (2008) Enrichment of a dioxin-dehalogenating *Dehalococcoides* species in two-liquid phase cultures. *Environ Microbiol* **10**: 2670-2683.
- Bunge, M., Kleikemper, J., Miniaci, C., Duc, L., Muusse, M.G., Hause, G., and Zeyer, J.** (2007) Benzoate-driven dehalogenation of chlorinated ethenes in microbial cultures from a contaminated aquifer. *Appl Microbiol Biotechnol* **76**: 1447-1456.
- Chard, B.K., Doucette, W.J., Chard, J.K., Bugbee, B., and Gorder, K.** (2006) Trichloroethylene uptake by apple and peach trees and transfer to fruit. *Environ Sci Technol* **40**: 4788-4793.
- Che, H., Bae, S., and Lee, W.** (2011) Degradation of trichloroethylene by Fenton reaction in pyrite suspension. *J Hazard Mater* **185**: 1355-1361.
- Chen, G.** (2004) Reductive dehalogenation of tetrachloroethylene by microorganisms: current knowledge and application strategies. *Appl Microbiol Biotechnol* **63**: 373-377.
- Chen, G., Hoag, G.E., Chedda, P., Nadim, F., Woody, B.A., and Dobbs, G.M.** (2001) The mechanism and applicability of in situ oxidation of trichloroethylene with Fenton's reagent. *J Hazard Mater* **87**: 171-186.
- Cheng, D., and He, J.** (2009) Isolation and characterization of "*Dehalococcoides*" sp. strain MB, which dechlorinates tetrachloroethene to trans-1,2-dichloroethene. *Appl Environ Microbiol* **75**: 5910-5918.
- Chou, C.C., Chen, C.H., Lee, T.T., and Peck, K.** (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res* **32**: e99.

- Chow, W.L., Cheng, D., Wang, S., and He, J.** (2010) Identification and transcriptional analysis of trans-DCE-producing reductive dehalogenases in Dehalococcoides species. *Isme J* **4**: 1020-1030.
- Christ, J.A., Ramsburg, C.A., Abriola, L.M., Pennell, K.D., and Löffler, F.E.** (2005) Coupling aggressive mass removal with microbial reductive dechlorination for remediation of DNAPL source zones: a review and assessment. *Environ Health Perspect* **113**: 465-477.
- Cichocka, D., Nikolausz, M., Haest, P.J., and Nijenhuis, I.** (2010) Tetrachloroethene conversion to ethene by a Dehalococcoides-containing enrichment culture from Bitterfeld. *FEMS Microbiol Ecol* **72**: 297-310.
- Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H.** (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* **4**: 265-270.
- Cochrane, G., Akhtar, R., Bonfield, J., Bower, L., Demiralp, F., Faruque, N. et al.** (2009) Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Res* **37**: D19-25.
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J. et al.** (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141-145.
- Coleman, N.V., and Spain, J.C.** (2003a) Epoxyalkane: coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of mycobacterium strain JS60. *J Bacteriol* **185**: 5536-5545.
- Coleman, N.V., and Spain, J.C.** (2003b) Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading Mycobacterium strains. *Appl Environ Microbiol* **69**: 6041-6046.
- Coleman, N.V., Mattes, T.E., Gossett, J.M., and Spain, J.C.** (2002a) Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. *Appl Environ Microbiol* **68**: 6162-6171.
- Coleman, N.V., Mattes, T.E., Gossett, J.M., and Spain, J.C.** (2002b) Biodegradation of cis-dichloroethene as the sole carbon source by a beta-proteobacterium. *Appl Environ Microbiol* **68**: 2726-2730.
- Collins, C., Laturnus, F., and Nepovim, A.** (2002) Remediation of BTEX and trichloroethene. Current knowledge with special emphasis on phytoremediation. *Environ Sci Pollut Res Int* **9**: 86-94.
- Contreras Lopez, M.C.** (2003) Determination of potentially bioaccumulating complex mixtures of organochlorine compounds in wastewater: a review. *Environment International* **28**: 751-759.
- Cruz-Martinez, K., Suttle, K.B., Brodie, E.L., Power, M.E., Andersen, G.L., and Banfield, J.F.** (2009) Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *Isme J* **3**: 738-744.
- Cupples, A.M.** (2008) Real-time PCR quantification of Dehalococcoides populations: methods and applications. *J Microbiol Methods* **72**: 1-11.
- Cupples, A.M., Spormann, A.M., and McCarty, P.L.** (2003) Growth of a Dehalococcoides-like microorganism on vinyl chloride and cis-dichloroethene as electron acceptors as determined by competitive PCR. *Appl Environ Microbiol* **69**: 953-959.
- Curtis, T.P., Sloan, W.T., and Scannell, J.W.** (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci U S A* **99**: 10494-10499.
- Cutter, L.A., Watts, J.E., Sowers, K.R., and May, H.D.** (2001) Identification of a microorganism that links its growth to the reductive dechlorination of 2,3,5,6-chlorobiphenyl. *Environ Microbiol* **3**: 699-709.

- Damborsky, J.** (1999) Tetrachloroethene-dehalogenating bacteria. *Folia Microbiol (Praha)* **44**: 247-262.
- Danko, A.S., Luo, M., Bagwell, C.E., Brigmon, R.L., and Freedman, D.L.** (2004) Involvement of linear plasmids in aerobic biodegradation of vinyl chloride. *Appl Environ Microbiol* **70**: 6092-6097.
- Daprato, R.C., Löffler, F.E., and Hughes, J.B.** (2007) Comparative Analysis of Three Tetrachloroethene to Ethene Halorespiring Consortia Suggests Functional Redundancy. *Environ Sci Technol* **41**: 2261-2269.
- Davis, G.B., Patterson, B.M., and Johnston, C.D.** (2009) Aerobic bioremediation of 1,2 dichloroethane and vinyl chloride at field scale. *J Contam Hydrol* **107**: 91-100.
- DeAngelis, K.M., Brodie, E.L., DeSantis, T.Z., Andersen, G.L., Lindow, S.E., and Firestone, M.K.** (2009) Selective progressive response of soil microbial community to wild oat roots. *Isme J* **3**: 168-178.
- Dekas, A.E., Poretsky, R.S., and Orphan, V.J.** (2009) Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* **326**: 422-426.
- DeLong, E.F., Wickham, G.S., and Pace, N.R.** (1989) Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243**: 1360-1363.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. et al.** (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069-5072.
- DeWeerd, K., Mandelco, L., Tanner, R., Woese, C., and Sufflita, J.** (1990) Desulfomonile tiedjei gen. nov. and sp. nov., a novel anaerobic, dehalogenating, sulfate-reducing bacterium. *Arch Microbiol* **154**: 23-30.
- DeWeerd, K.A., Flanagan, W.P., Brennan, M.J., Principe, J.M., and Spivack, J.L.** (1998) Biodegradation of Trichloroethylene and Dichloromethane in Contaminated Soil and Groundwater. *Bioremediation Journal* **2**: 29-42.
- Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M. et al.** (2008) Functional metagenomic profiling of nine biomes. *Nature* **452**: 629-632.
- Dong, Y., Butler, E.C., Philp, R.P., and Krumholz, L.R.** (2011) Impacts of microbial community composition on isotope fractionation during reductive dechlorination of tetrachloroethylene. *Biodegradation* **22**: 431-444.
- Doucette, W.J., Chard, J.K., Fabrizius, H., Crouch, C., Petersen, M.R., Carlsen, T.E. et al.** (2007) Trichloroethylene Uptake into Fruits and Vegetables: Three-Year Field Monitoring Study. *Environ Sci Technol* **41**: 2505-2509.
- Drzyzga, O., and Gottschal, J.C.** (2002) Tetrachloroethene dehalorespiration and growth of Desulfitobacterium frappieri TCE1 in strict dependence on the activity of Desulfovibrio fructosivorans. *Appl Environ Microbiol* **68**: 642-649.
- Dua, M., Singh, A., Sethunathan, N., and Johri, A.K.** (2002) Biotechnology and bioremediation: successes and limitations. *Appl Microbiol Biotechnol* **59**: 143-152.
- Duc, L., Neuenschwander, S., Rehrauer, H., Wagner, U., Sobek, J., Schlappbach, R., and Zeyer, J.** (2009) Development and experimental validation of a nifH oligonucleotide microarray to study diazotrophic communities in a glacier forefield. *Environ Microbiol* **11**: 2179-2189.
- Dufva, M.** (2005) Fabrication of high quality microarrays. *Biomol Eng* **22**: 173-184.
- Dugat-Bony, E., Missaoui, M., Peyretailade, E., Biderre-Petit, C., Bouzid, O., Gouinaud, C. et al.** (2011) HiSpOD: probe design for functional DNA microarrays. *Bioinformatics* **27**: 641-648.
- Duhamel, M., and Edwards, E.A.** (2007) Growth and Yields of Dechlorinators, Acetogens, and Methanogens during Reductive Dechlorination of Chlorinated Ethenes and Dihaloelimination of 1,2-Dichloroethane. *Environ Sci Technol* **41**: 2303-2310.

- Duhamel, M., Mo, K., and Edwards, E.A.** (2004) Characterization of a Highly Enriched Dehalococcoides-Containing Culture That Grows on Vinyl Chloride and Trichloroethene. *Appl Environ Microbiol* **70**: 5538-5545.
- Duhamel, M., Wehr, S.D., Yu, L., Rizvi, H., Seepersad, D., Dworatzek, S. et al.** (2002) Comparison of anaerobic dechlorinating enrichment cultures maintained on tetrachloroethene, trichloroethene, cis-dichloroethene and vinyl chloride. *Water Res* **36**: 4193-4202.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G. et al.** (2009) Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133-138.
- El Fantroussi, S., Naveau, H., and Agathos, S.N.** (1998) Anaerobic Dechlorinating Bacteria. *Biotechnology Progress* **14**: 167-188.
- Elango, V.K., Ligenstoffer, A.S., and Fathepure, B.Z.** (2006) Biodegradation of vinyl chloride and cis-dichloroethene by a *Ralstonia* sp. strain TRW-1. *Appl Microbiol Biotechnol* **72**: 1270-1275.
- Ellis, D.E., Lutz, E.J., Odom, J.M., Buchanan, R.J., Bartlett, C.L., Lee, M.D. et al.** (2000) Bioaugmentation for Accelerated In Situ Anaerobic Bioremediation. *Environ Sci Technol* **34**: 2254-2260.
- Ely, R.L., Williamson, K.J., Hyman, M.R., and Arp, D.J.** (1997) Cometabolism of chlorinated solvents by nitrifying bacteria: kinetics, substrate interactions, toxicity effects, and bacterial response. *Biotechnol Bioeng* **54**: 520-534.
- Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., and Hugenholtz, P.** (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *Isme J* **4**: 642-647.
- Ensign, S.A., Hyman, M.R., and Arp, D.J.** (1992) Cometabolic degradation of chlorinated alkenes by alkene monooxygenase in a propylene-grown *Xanthobacter* strain. *Appl Environ Microbiol* **58**: 3038-3046.
- EPA** (2000) Engineered Approaches to In Situ Bioremediation of Chlorinated Solvents: Fundamentals and Field Applications.
- Epelde, L., Becerril, J.M., Kowalchuk, G.A., Deng, Y., Zhou, J., and Garbisu, C.** (2010) Impact of metal pollution and *Thlaspi caerulescens* growth on soil microbial communities. *Appl Environ Microbiol* **76**: 7843-7853.
- Fagervold, S.K., May, H.D., and Sowers, K.R.** (2007) Microbial reductive dechlorination of aroclor 1260 in Baltimore harbor sediment microcosms is catalyzed by three phylotypes within the phylum Chloroflexi. *Appl Environ Microbiol* **73**: 3009-3018.
- Fagervold, S.K., Watts, J.E., May, H.D., and Sowers, K.R.** (2005) Sequential reductive dechlorination of meta-chlorinated polychlorinated biphenyl congeners in sediment microcosms by two different Chloroflexi phylotypes. *Appl Environ Microbiol* **71**: 8085-8090.
- Fan, S., and Scow, K.M.** (1993) Biodegradation of trichloroethylene and toluene by indigenous microbial populations in soil. *Appl Environ Microbiol* **59**: 1911-1918.
- Fathepure, B.Z., and Boyd, S.A.** (1988) Dependence of tetrachloroethylene dechlorination on methanogenic substrate consumption by *Methanosarcina* sp. strain DCM. *Appl Environ Microbiol* **54**: 2976-2980.
- Fathepure, B.Z., Elango, V.K., Singh, H., and Bruner, M.A.** (2005) Bioaugmentation potential of a vinyl chloride-assimilating *Mycobacterium* sp., isolated from a chloroethene-contaminated aquifer. *FEMS Microbiol Lett* **248**: 227-234.
- Fetzner, S.** (1998) Bacterial dehalogenation. *Appl Microbiol Biotechnol* **50**: 633-657.
- Field, J.A., and Sierra-Alvarez, R.** (2008) Microbial transformation and degradation of polychlorinated biphenyls. *Environ Pollut* **155**: 1-12.
- Finneran, K.T., Forbush, H.M., VanPraagh, C.V., and Lovley, D.R.** (2002) *Desulfitobacterium metallireducens* sp. nov., an anaerobic bacterium that couples growth to

the reduction of metals and humic acids as well as chlorinated compounds. *Int J Syst Evol Microbiol* **52**: 1929-1935.

Fletcher, K.E., Nijenhuis, I., Richnow, H.H., and Löffler, F.E. (2011) Stable carbon isotope enrichment factors for cis-1,2-dichloroethene and vinyl chloride reductive dechlorination by *Dehalococcoides*. *Environ Sci Technol* **45**: 2951-2957.

Franke-Whittle, I.H., Goberna, M., Pfister, V., and Insam, H. (2009) Design and development of the ANAEROCHIP microarray for investigation of methanogenic communities. *J Microbiol Methods* **79**: 279-288.

Friis, A.K., Kofoed, J.L., Heron, G., Albrechtsen, H.J., and Bjerg, P.L. (2007) Microcosm evaluation of bioaugmentation after field-scale thermal treatment of a TCE-contaminated aquifer. *Biodegradation* **18**: 661-674.

Fung, J.M., Morris, R.M., Adrian, L., and Zinder, S.H. (2007) Expression of reductive dehalogenase genes in *Dehalococcoides ethenogenes* strain 195 growing on tetrachloroethene, trichloroethene, or 2,3-dichlorophenol. *Appl Environ Microbiol* **73**: 4439-4445.

Furukawa, K. (2003) 'Super bugs' for bioremediation. *Trends Biotechnol* **21**: 187-190.

Furukawa, K. (2006) Oxygenases and dehalogenases: molecular approaches to efficient degradation of chlorinated environmental pollutants. *Biosci Biotechnol Biochem* **70**: 2335-2348.

Futagami, T., Goto, M., and Furukawa, K. (2008) Biochemical and genetic bases of dehalorespiration. *Chem Rec* **8**: 1-12.

Futagami, T., Morono, Y., Terada, T., Kaksonen, A.H., and Inagaki, F. (2009) Dehalogenation Activities and Distribution of Reductive Dehalogenase Homologous Genes in Marine Subsurface Sediments. *Appl Environ Microbiol* **75**: 6905-6909.

Futamata, H., Kaiya, S., Sugawara, M., and Hiraishi, A. (2009) Phylogenetic and Transcriptional Analyses of a Tetrachloroethene-Dechlorinating "*Dehalococcoides*" Enrichment Culture TUT2264 and Its Reductive-Dehalogenase Genes. *Microbes Environ* **24**: 330-337.

Gans, J., Wolinsky, M., and Dunbar, J. (2005) Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. *Science* **309**: 1387-1390.

Gao, Y. (2002) Glycopeptide antibiotics and development of inhibitors to overcome vancomycin resistance. *Natural Product Reports* **19**: 100-107.

Gavaskar, A.R. (1999) Design and construction techniques for permeable reactive barriers. *J Hazard Mater* **68**: 41-71.

Gentry, T.J., Wickham, G.S., Schadt, C.W., He, Z., and Zhou, J. (2006) Microarray applications in microbial ecology research. *Microb Ecol* **52**: 159-175.

Gerritse, J., Renard, V., Pedro Gomes, T.M., Lawson, P.A., Collins, M.D., and Gottschal, J.C. (1996) *Desulfitobacterium* sp. strain PCE1, an anaerobic bacterium that can grow by reductive dechlorination of tetrachloroethene or ortho-chlorinated phenols. *Arch Microbiol* **165**: 132-140.

Gerritse, J., Drzyzga, O., Kloetstra, G., Keijmel, M., Wiersum, L.P., Hutson, R. et al. (1999) Influence of different electron donors and acceptors on dehalorespiration of tetrachloroethene by *Desulfitobacterium frappieri* TCE1. *Appl Environ Microbiol* **65**: 5212-5221.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A. et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52-56.

Goltz, M.N., Gandhi, R.K., Gorelick, S.M., Hopkins, G.D., Smith, L.H., Timmins, B.H., and McCarty, P.L. (2005) Field evaluation of in situ source reduction of trichloroethylene in groundwater using bioenhanced in-well vapor stripping. *Environ Sci Technol* **39**: 8963-8970.

- Gossett, J.M.** (2010) Sustained aerobic oxidation of vinyl chloride at low oxygen concentrations. *Environ Sci Technol* **44**: 1405-1411.
- Grandel, S., and Dahmke, A.** (2004) Monitored natural attenuation of chlorinated solvents: assessment of potential and limitations. *Biodegradation* **15**: 371-386.
- Gribble, G.W.** (1994) The abundant natural sources and uses of chlorinated chemicals. *Am J Public Health* **84**: 1183.
- Gribble, G.W.** (2003) The diversity of naturally produced organohalogenes. *Chemosphere* **52**: 289-297.
- Gribble, G.W.** (2004) Amazing Organohalogenes. *Eurochlor*.
- Groster, A., and Edwards, E.A.** (2006) A 1,1,1-trichloroethane-degrading anaerobic mixed microbial culture enhances biotransformation of mixtures of chlorinated ethenes and ethanes. *Appl Environ Microbiol* **72**: 7849-7856.
- Gudbjerg, J., Sonnenborg, T.O., and Jensen, K.H.** (2004) Remediation of NAPL below the water table by steam-induced heat conduction. *J Contam Hydrol* **72**: 207-225.
- Guerrero, R., and Berlanga, M.** (2006) Life's unity and flexibility: the ecological link. *Int Microbiol* **9**: 225-235.
- Guiot, S.R., Cimpoea, R., Kuhn, R., and Alaplantive, A.** (2008) Electrolytic methanogenic-methanotrophic coupling for tetrachloroethylene bioremediation: proof of concept. *Environ Sci Technol* **42**: 3011-3017.
- Guschin, D.Y., Mobarry, B.K., Proudnikov, D., Stahl, D.A., Rittmann, B.E., and Mirzabekov, A.D.** (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl Environ Microbiol* **63**: 2397-2402.
- Hamamura, N., Page, C., Long, T., Semprini, L., and Arp, D.J.** (1997) Chloroform Cometabolism by Butane-Grown CF8, *Pseudomonas butanovora*, and *Mycobacterium vaccae* JOB5 and Methane-Grown *Methylosinus trichosporium* OB3b. *Appl Environ Microbiol* **63**: 3607-3613.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M.** (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: R245-249.
- Harker, A.R., and Kim, Y.** (1990) Trichloroethylene degradation by two independent aromatic-degrading pathways in *Alcaligenes eutrophus* JMP134. *Appl Environ Microbiol* **56**: 1179-1181.
- Harms, H., Schlosser, D., and Wick, L.Y.** (2011) Untapped potential: exploiting fungi in bioremediation of hazardous chemicals. *Nat Rev Microbiol* **9**: 177-192.
- Hartmans, S., and De Bont, J.A.** (1992) Aerobic vinyl chloride metabolism in *Mycobacterium aurum* L1. *Appl Environ Microbiol* **58**: 1220-1226.
- Hartmans, S., de Bont, J.A.M., Tramper, J., and Luyben, K.C.A.M.** (1985) Bacterial degradation of vinyl chloride. *Biotechnology Letters* **7**: 383-388.
- Hata, J., Takamizawa, K., Miyata, N., and Iwahori, K.** (2003) Biodegradation of cis-1,2-dichloroethylene and vinyl chloride in anaerobic cultures enriched from landfill leachate sediment under Fe(III)-reducing conditions. *Biodegradation* **14**: 275-283.
- Hata, J., Miyata, N., Kim, E.S., Takamizawa, K., and Iwahori, K.** (2004) Anaerobic degradation of cis-1,2-dichloroethylene and vinyl chloride by *Clostridium* sp. strain DC1 isolated from landfill leachate sediment. *J Biosci Bioeng* **97**: 196-201.
- He, J., Robrock, K.R., and Alvarez-Cohen, L.** (2006) Microbial reductive debromination of polybrominated diphenyl ethers (PBDEs). *Environ Sci Technol* **40**: 4429-4434.
- He, J., Holmes, V.F., Lee, P.K., and Alvarez-Cohen, L.** (2007a) Influence of vitamin B12 and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Appl Environ Microbiol* **73**: 2847-2853.

- He, J., Ritalahti, K.M., Yang, K.L., Koenigsberg, S.S., and Löffler, F.E.** (2003) Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* **424**: 62-65.
- He, J., Sung, Y., Krajmalnik-Brown, R., Ritalahti, K.M., and Löffler, F.E.** (2005) Isolation and characterization of Dehalococcoides sp. strain FL2, a trichloroethene (TCE)- and 1,2-dichloroethene-respiring anaerobe. *Environ Microbiol* **7**: 1442-1450.
- He, J., Sung, Y., Dollhopf, M.E., Fathepure, B.Z., Tiedje, J.M., and Löffler, F.E.** (2002) Acetate versus Hydrogen as Direct Electron Donors To Stimulate the Microbial Reductive Dechlorination Process at Chloroethene-Contaminated Sites. *Environ Sci Technol* **36**: 3945-3952.
- He, Z., Van Nostrand, J.D., Deng, Y., and Zhou, J.Z.** (2011) Development and applications of functional gene microarrays in the analysis of the functional diversity, composition, and structure of microbial communities. *Front Environ Sci Engin China* **5**: 1-20.
- He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C. et al.** (2007b) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *Isme J* **1**: 67-77.
- He, Z., Deng, Y., Van Nostrand, J.D., Tu, Q., Xu, M., Hemme, C.L. et al.** (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *Isme J* **4**: 1167-1179.
- He, Z.L., Van Nostrand, J.D., Wu, L.Y., and Zhou, J.Z.** (2008) Development and application of functional gene arrays for microbial community analysis. *Trans Nonferrous Met Soc China* **18**: 1319-1327.
- Heald, S., and Jenkins, R.O.** (1994) *Trichloroethylene removal and oxidation toxicity mediated by toluene dioxygenase of Pseudomonas putida*. Washington, DC, ETATS-UNIS: American Society for Microbiology.
- Heimann, A., Jakobsen, R., and Blodau, C.** (2009) Energetic constraints on H₂-dependent terminal electron accepting processes in anoxic environments: a review of observations and model approaches. *Environ Sci Technol* **44**: 24-33.
- Heimann, A.C., Batstone, D.J., and Jakobsen, R.** (2006) Methanosarcina spp. drive vinyl chloride dechlorination via interspecies hydrogen transfer. *Appl Environ Microbiol* **72**: 2942-2949.
- Hemme, C.L., Deng, Y., Gentry, T.J., Fields, M.W., Wu, L., Barua, S. et al.** (2010) Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *Isme J* **4**: 660-672.
- Hendrickson, E.R., Payne, J.A., Young, R.M., Starr, M.G., Perry, M.P., Fahnestock, S. et al.** (2002) Molecular analysis of Dehalococcoides 16S ribosomal DNA from chloroethene-contaminated sites throughout North America and Europe. *Appl Environ Microbiol* **68**: 485-495.
- Henry, S.M., Hardcastle, C.H., and Warner, S.D.** (2002) Chlorinated Solvent and DNAPL Remediation: An Overview of Physical, Chemical, and Biological Processes. In *Chlorinated Solvent and DNAPL Remediation*: American Chemical Society, pp. 1-20.
- Hiraishi, A.** (2008) Biodiversity of dehalorespiring bacteria with special emphasis on polychlorinated biphenyl/dioxin dechlorinators. *Microbes Environ* **23**: 1-12.
- Hiraishi, A., Sakamaki, N., Miyakoda, H., Maruyama, T., Kato, K., and Futamata, H.** (2005) Estimation of "Dehalococcoides" Populations in Lake Sediment Contaminated with Low Levels of Polychlorinated Dioxins. *Microbes Environ* **20**: 216-226.
- Holliger, C., and Schraa, G.** (1994) Physiological meaning and potential for application of reductive dechlorination by anaerobic bacteria. *FEMS Microbiol Rev* **15**: 297-305.
- Holliger, C., Wohlfarth, G., and Diekert, G.** (1998a) Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiol Rev* **22**: 383-398.

- Holliger, C., Hahn, D., Harmsen, H., Ludwig, W., Schumacher, W., Tindall, B. et al.** (1998b) *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetra- and trichloroethene in an anaerobic respiration. *Arch Microbiol* **169**: 313-321.
- Holmes, V.F., He, J., Lee, P.K., and Alvarez-Cohen, L.** (2006) Discrimination of multiple *Dehalococcoides* strains in a trichloroethene enrichment by quantification of their reductive dehalogenase genes. *Appl Environ Microbiol* **72**: 5877-5883.
- Holscher, T., Krajmalnik-Brown, R., Ritalahti, K.M., Von Wintzingerode, F., Gorisch, H., Löffler, F.E., and Adrian, L.** (2004) Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*. *Appl Environ Microbiol* **70**: 5290-5297.
- Hugenholtz, P., and Stackebrandt, E.** (2004) Reclassification of *Sphaerobacter thermophilus* from the subclass *Sphaerobacteridae* in the phylum *Actinobacteria* to the class *Thermomicrobia* (emended description) in the phylum *Chloroflexi* (emended description). *Int J Syst Evol Microbiol* **54**: 2049-2051.
- Huyghe, A., Francois, P., Charbonnier, Y., Tangomo-Bento, M., Bonetti, E.J., Paster, B.J. et al.** (2008) Novel microarray design strategy to study complex bacterial communities. *Appl Environ Microbiol* **74**: 1876-1885.
- Hyman, M.R., Russell, S.A., Ely, R.L., Williamson, K.J., and Arp, D.J.** (1995) Inhibition, Inactivation, and Recovery of Ammonia-Oxidizing Activity in Cometabolism of Trichloroethylene by *Nitrosomonas europaea*. *Appl Environ Microbiol* **61**: 1480-1487.
- Iwai, S., Kurisu, F., Urakawa, H., Yagi, O., and Furumai, H.** (2007) Development of a 60-mer oligonucleotide microarray on the basis of benzene monooxygenase gene diversity. *Appl Microbiol Biotechnol* **75**: 929-939.
- Iwai, S., Kurisu, F., Urakawa, H., Yagi, O., Kasuga, I., and Furumai, H.** (2008) Development of an oligonucleotide microarray to detect di- and monooxygenase genes for benzene degradation in soil. *FEMS Microbiol Lett* **285**: 111-121.
- Jablonski, P.E., and Ferry, J.G.** (1992) Reductive dechlorination of trichloroethylene by the CO-reduced CO dehydrogenase enzyme complex from *Methanosarcina thermophila*. *FEMS Microbiol Lett* **75**: 55-59.
- Jackson, R.M.T.J.R.F.W.G., Jackson, R.M.T.J.R.F.W.G., and Jackson, R.M.T.J.R.F.W.G.** (1973) *Thermomicrobium*, a New Genus of Extremely Thermophilic Bacteria. *Int J Syst Bacteriol* **23**: 28-36.
- James, C.A., Xin, G., Doty, S.L., Muiznieks, I., Newman, L., and Strand, S.E.** (2009) A mass balance study of the phytoremediation of perchloroethylene-contaminated groundwater. *Environ Pollut* **157**: 2564-2569.
- Janssen, D.B., Dinkla, I.J., Poelarends, G.J., and Terpstra, P.** (2005) Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ Microbiol* **7**: 1868-1882.
- Jawitz, J.W., Sillan, R.K., Annable, M.D., Rao, P.S.C., and Warner, K.** (2000) In-Situ Alcohol Flushing of a DNAPL Source Zone at a Dry Cleaner Site. *Environ Sci Technol* **34**: 3722-3729.
- Jennings, L.K., Chartrand, M.M., Lacrampe-Couloume, G., Lollar, B.S., Spain, J.C., and Gossett, J.M.** (2009) Proteomic and transcriptomic analyses reveal genes upregulated by cis-dichloroethene in *Polaromonas* sp. strain JS666. *Appl Environ Microbiol* **75**: 3733-3744.
- Jin, Y.O., and Mattes, T.E.** (2008) Adaptation of aerobic, ethene-assimilating *Mycobacterium* strains to vinyl chloride as a growth substrate. *Environ Sci Technol* **42**: 4784-4789.
- Jin, Y.O., and Mattes, T.E.** (2010) A quantitative PCR assay for aerobic, vinyl chloride- and ethene-assimilating microorganisms in groundwater. *Environ Sci Technol* **44**: 9036-9041.

- John, M., Schmitz, R.P., Westermann, M., Richter, W., and Diekert, G.** (2006) Growth substrate dependent localization of tetrachloroethene reductive dehalogenase in *Sulfurospirillum multivorans*. *Arch Microbiol* **186**: 99-106.
- Johnson, D.R., Lee, P.K., Holmes, V.F., and Alvarez-Cohen, L.** (2005) An internal reference technique for accurately quantifying specific mRNAs by real-time PCR with application to the *tceA* reductive dehalogenase gene. *Appl Environ Microbiol* **71**: 3866-3871.
- Jordan, A.** (2003) Volcanic Formation of Halogenated Organic Compounds. In *The Handbook of Environmental Chemistry*: Springer-Verlag Berlin Heidelberg, pp. 121-139.
- Jordan, A., Harnisch, J., Borchers, R., Le Guern, F., and Shinohara, H.** (2000) Volcanogenic Halocarbons. *Environ Sci Technol* **34**: 1122-1124.
- Kalin, R.M.** (2004) Engineered passive bioreactive barriers: risk-managing the legacy of industrial soil and groundwater pollution. *Curr Opin Microbiol* **7**: 227-238.
- Kao, C.M., Chen, S.C., and Liu, J.K.** (2001) Development of a biobarrier for the remediation of PCE-contaminated aquifer. *Chemosphere* **43**: 1071-1078.
- Kao, C.M., Chen, Y.L., Chen, S.C., Yeh, T.Y., and Wu, W.S.** (2003) Enhanced PCE dechlorination by biobarrier systems under different redox conditions. *Water Res* **37**: 4885-4894.
- Kao, C.M., Huang, K.D., Wang, J.Y., Chen, T.Y., and Chien, H.Y.** (2008) Application of potassium permanganate as an oxidant for in situ oxidation of trichloroethylene-contaminated groundwater: A laboratory and kinetics study. *J Hazard Mater* **153**: 919-927.
- Kaslusky, S.F., and Udell, K.S.** (2002) A theoretical model of air and steam co-injection to prevent the downward migration of DNAPLs during steam-enhanced extraction. *J Contam Hydrol* **55**: 213-232.
- Kawasaki, E.S.** (2006) The end of the microarray Tower of Babel: will universal standards lead the way? *J Biomol Tech* **17**: 200-206.
- Kelly, J.J., Siripong, S., McCormack, J., Janus, L.R., Urakawa, H., El Fantroussi, S. et al.** (2005) DNA microarray detection of nitrifying bacterial 16S rRNA in wastewater treatment plant samples. *Water Res* **39**: 3229-3238.
- Keppen, O.I., Tourova, T.P., Kuznetsov, B.B., Ivanovsky, R.N., and Gorlenko, V.M.** (2000) Proposal of Oscillochloridaceae fam. nov. on the basis of a phylogenetic analysis of the filamentous anoxygenic phototrophic bacteria, and emended description of Oscillochloris and Oscillochloris trichoides in comparison with further new isolates. *Int J Syst Evol Microbiol* **50 Pt 4**: 1529-1537.
- Khan, F.I., Husain, T., and Hejazi, R.** (2004) An overview and analysis of site remediation technologies. *J Environ Manage* **71**: 95-122.
- Kim, E.-S., Nomura, I., Hasegawa, Y., and Takamizawa, K.** (2006) Characterization of a newly isolated cis-1,2-dichloroethylene and aliphatic compound-degrading bacterium, *Clostridium* sp. strain KYT-1. *Biotechnol Bioprocess Eng* **11**: 553-556.
- Kim, Y., Istok, J.D., and Semprini, L.** (2008) Single-well, gas-sparging tests for evaluating the in situ aerobic cometabolism of cis-1,2-dichloroethene and trichloroethene. *Chemosphere* **71**: 1654-1664.
- Kittelmann, S., and Friedrich, M.W.** (2008) Novel uncultured Chloroflexi dechlorinate perchloroethene to trans-dichloroethene in tidal flat sediments. *Environ Microbiol* **10**: 1557-1570.
- Knox, R.C., Sabatini, D.A., Harwell, J.H., Brown, R.E., West, C.C., Blaha, F., and Griffin, C.** (1997) Surfactant Remediation Field Demonstration Using a Vertical Circulation Well. *Ground Water* **35**: 948-953.
- Koh, S.C., Bowman, J.P., and Sayler, G.S.** (1993) Soluble Methane Monooxygenase Production and Trichloroethylene Degradation by a Type I Methanotroph, *Methylomonas methanica* 68-1. *Appl Environ Microbiol* **59**: 960-967.

- Kohlhoff, K.J., Sosnick, M.H., Hsu, W.T., Pande, V.S., and Altman, R.B.** (2011) CAMPAIGN: an open-source library of GPU-accelerated data clustering algorithms. *Bioinformatics* **27**: 2322-2323.
- Korlach, J., Bjornson, K.P., Chaudhuri, B.P., Cicero, R.L., Flusberg, B.A., Gray, J.J. et al.** (2010) Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* **472**: 431-455.
- Krajmalnik-Brown, R., Sung, Y., Ritalahti, K.M., Saunders, F.M., and Löffler, F.E.** (2007) Environmental distribution of the trichloroethene reductive dehalogenase gene (*tceA*) suggests lateral gene transfer among Dehalococcoides. *FEMS Microbiol Ecol* **59**: 206-214.
- Krajmalnik-Brown, R., Holscher, T., Thomson, I.N., Saunders, F.M., Ritalahti, K.M., and Löffler, F.E.** (2004) Genetic identification of a putative vinyl chloride reductase in Dehalococcoides sp. strain BAV1. *Appl Environ Microbiol* **70**: 6347-6351.
- Kramer, U.** (2005) Phytoremediation: novel approaches to cleaning up polluted soils. *Curr Opin Biotechnol* **16**: 133-141.
- Krauter, P., MacQueen, D., Horn, J., and Bishop, D.** (1995) *Effect of subsurface electrical heating and steam injection on the indigenous microbial community*: Lawrence Livermore National Laboratory.
- Krumholz, L.R.** (1997) Desulfuromonas chloroethenica sp. nov. Uses Tetrachloroethylene and Trichloroethylene as Electron Acceptors. *Int J Syst Bacteriol* **47**: 1262-1263.
- Kube, M., Beck, A., Zinder, S.H., Kuhl, H., Reinhardt, R., and Adrian, L.** (2005) Genome sequence of the chlorinated compound-respiring bacterium Dehalococcoides species strain CBDB1. *Nat Biotechnol* **23**: 1269-1273.
- Kyselkova, M., Kopecky, J., Felfoldi, T., Cermak, L., Omelka, M., Grundmann, G.L. et al.** (2008) Development of a 16S rRNA gene-based prototype microarray for the detection of selected actinomycetes genera. *Antonie Van Leeuwenhoek* **94**: 439-453.
- Landa, A.S., Sipkema, E.M., Weijma, J., Beenackers, A.A., Dolfing, J., and Janssen, D.B.** (1994) Cometabolic degradation of trichloroethylene by Pseudomonas cepacia G4 in a chemostat with toluene as the primary substrate. *Appl Environ Microbiol* **60**: 3368-3374.
- Lash, L.H., and Parker, J.C.** (2001) Hepatic and renal toxicities associated with perchloroethylene. *Pharmacol Rev* **53**: 177-208.
- Laternus, F.** (2001) Marine Macroalgae in Polar Regions as Natural Sources for Volatile Organohalogenes. *Environ Sci Pollut Res Int* **2**: 103-108.
- Leahy, J.G., Byrne, A.M., and Olsen, R.H.** (1996) Comparison of factors influencing trichloroethylene degradation by toluene-oxidizing bacteria. *Appl Environ Microbiol* **62**: 825-833.
- Lee, Chi, Y., Chan, Yu, C., Lin, and Chin, L.** (2005) The effect of salinity on trichloroethylene co-metabolism by mixed cultures enriched on phenol. *World J Microbiol Biotechnol* **21**: 7.
- Lee, C.Y., and Liu, W.D.** (2006) The effect of salinity conditions on kinetics of trichloroethylene biodegradation by toluene-oxidizing cultures. *J Hazard Mater* **137**: 541-549.
- Lee, I.S., Bae, J.H., and McCarty, P.L.** (2007) Comparison between acetate and hydrogen as electron donors and implications for the reductive dehalogenation of PCE and TCE. *J Contam Hydrol* **94**: 76-85.
- Lee, N., Nielsen, P.H., Andreasen, K.H., Juretschko, S., Nielsen, J.L., Schleifer, K.H., and Wagner, M.** (1999) Combination of fluorescent in situ hybridization and microautoradiography-a new tool for structure-function analyses in microbial ecology. *Appl Environ Microbiol* **65**: 1289-1297.
- Lee, P.K., Cheng, D., Hu, P., West, K.A., Dick, G.J., Brodie, E.L. et al.** (2011) Comparative genomics of two newly isolated Dehalococcoides strains and an enrichment using a genus microarray. *Isme J* **5**: 1014-1024.

- Leigh, M.B., Pellizari, V.H., Uhlik, O., Sutka, R., Rodrigues, J., Ostrom, N.E. et al.** (2007) Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *Isme J* **1**: 134-148.
- Leininger, S., Urich, T., Schlöter, M., Schwark, L., Qi, J., Nicol, G.W. et al.** (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806-809.
- Lemoine, S., Combes, F., and Le Crom, S.** (2009) An evaluation of custom microarray applications: the oligonucleotide design challenge. *Nucleic Acids Res* **37**: 1726-1739.
- Lenczewski, M., Jardine, P., McKay, L., and Layton, A.** (2003) Natural attenuation of trichloroethylene in fractured shale bedrock. *J Contam Hydrol* **64**: 151-168.
- Li, T., Wu, T.D., Mazeas, L., Toffin, L., Guerquin-Kern, J.L., Leblon, G., and Bouchez, T.** (2008) Simultaneous analysis of microbial identity and function using NanoSIMS. *Environ Microbiol* **10**: 580-588.
- Li, Z., and Hanlie, H.** (2008) Combination of surfactant solubilization with permanganate oxidation for DNAPL remediation. *Water Res* **42**: 605-614.
- Liang, C., Lee, I.L., Hsu, I.Y., Liang, C.-P., and Lin, Y.-L.** (2008) Persulfate oxidation of trichloroethylene with and without iron activation in porous media. *Chemosphere* **70**: 426-435.
- Liang, Y., Li, G., Nostrand, J.D.V., He, Z., Wu, L., Deng, Y. et al.** (2009) Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field. *FEMS Microbiol Ecol* **70**: 324-333.
- Liang, Y., Van Nostrand, J.D., Deng, Y., He, Z., Wu, L., Zhang, X. et al.** (2011) Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *Isme J* **5**: 403-413.
- Liles, M.R., Turkmen, O., Manske, B.F., Zhang, M., Rouillard, J.-M., George, I. et al.** (2010) A phylogenetic microarray targeting 16S rRNA genes from the bacterial division Acidobacteria reveals a lineage-specific distribution in a soil clay fraction. *Soil Biol Biochem* **42**: 739-747.
- Lin, Y., Gao, F., and Zhang, C.T.** (2010) Functionality of essential genes drives gene strand-bias in bacterial genomes. *Biochem Biophys Res Commun* **396**: 472-476.
- Little, C.D., Palumbo, A.V., Herbes, S.E., Lidstrom, M.E., Tyndall, R.L., and Gilmer, P.J.** (1988) Trichloroethylene biodegradation by a methane-oxidizing bacterium. *Appl Environ Microbiol* **54**: 951-956.
- Lock, E.A., and Reed, C.J.** (2006) Trichloroethylene: mechanisms of renal toxicity and renal cancer and relevance to risk assessment. *Toxicol Sci* **91**: 313-331.
- Löffler, F.E., and Edwards, E.A.** (2006) Harnessing microbial activities for environmental cleanup. *Curr Opin Biotechnol* **17**: 274-284.
- Lohner, S.T., Becker, D., Mangold, K.M., and Tiehm, A.** (2011) Sequential reductive and oxidative biodegradation of chloroethenes stimulated in a coupled bioelectro-process. *Environ Sci Technol* **45**: 6491-6497.
- Long, C.M., and Borden, R.C.** (2006) Enhanced reductive dechlorination in columns treated with edible oil emulsion. *J Contam Hydrol* **87**: 54-72.
- Lontoh, S., and Semrau, J.D.** (1998) Methane and Trichloroethylene Degradation by *Methylosinus trichosporium* OB3b Expressing Particulate Methane Monooxygenase. *Appl Environ Microbiol* **64**: 1106-1114.
- Lovley, D.R.** (2003) Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Microbiol* **1**: 35-44.
- Loy, A., and Bodrossy, L.** (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clin Chim Acta* **363**: 106-119.

- Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J. et al.** (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68**: 5064-5081.
- Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoecker, K., Baranyi, C. et al.** (2005) 16S rRNA Gene-Based Oligonucleotide Microarray for Environmental Monitoring of the Betaproteobacterial Order "Rhodocyclales". *Appl Environ Microbiol* **71**: 1373-1386.
- Lueders, T., and Friedrich, M.W.** (2003) Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Appl Environ Microbiol* **69**: 320-326.
- Lueders, T., Pommerenke, B., and Friedrich, M.W.** (2004) Stable-isotope probing of microorganisms thriving at thermodynamic limits: syntrophic propionate oxidation in flooded soil. *Appl Environ Microbiol* **70**: 5778-5786.
- Luijten, M.L., Roelofsen, W., Langenhoff, A.A., Schraa, G., and Stams, A.J.** (2004) Hydrogen threshold concentrations in pure cultures of halo-respiring bacteria and at a site polluted with chlorinated ethenes. *Environ Microbiol* **6**: 646-650.
- Luijten, M.L., de Weert, J., Smidt, H., Boschker, H.T., de Vos, W.M., Schraa, G., and Stams, A.J.** (2003) Description of *Sulfurospirillum halo-respirans* sp. nov., an anaerobic, tetrachloroethene-respiring bacterium, and transfer of *Dehalospirillum multivorans* to the genus *Sulfurospirillum* as *Sulfurospirillum multivorans* comb. nov. *Int J Syst Evol Microbiol* **53**: 787-793.
- Ma, X., and Burken, J.G.** (2003) TCE diffusion to the atmosphere in phytoremediation applications. *Environ Sci Technol* **37**: 2534-2539.
- MacDonald, J.A., and Kavanaugh, M.C.** (1994) Restoring contaminated groundwater: an achievable goal? *Environ Sci Technol* **28**: 362A-368A.
- Mackay, D.M., and Cherry, J.A.** (1989) Groundwater contamination: pump-and-treat remediation. *Environ Sci Technol* **23**: 630-636.
- Magnuson, J.K., Romine, M.F., Burris, D.R., and Kingsley, M.T.** (2000) Trichloroethene reductive dehalogenase from *Dehalococcoides ethenogenes*: sequence of tceA and substrate range characterization. *Appl Environ Microbiol* **66**: 5141-5147.
- Magnuson, J.K., Stern, R.V., Gossett, J.M., Zinder, S.H., and Burris, D.R.** (1998) Reductive dechlorination of tetrachloroethene to ethene by a two-component enzyme pathway. *Appl Environ Microbiol* **64**: 1270-1275.
- Maillard, J., Schumacher, W., Vazquez, F., Regeard, C., Hagen, W.R., and Holliger, C.** (2003) Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus*. *Appl Environ Microbiol* **69**: 4628-4638.
- Maillard, J., Charnay, M.-P., Regeard, C., Rohrbach-Brandt, E., Rouzeau-Szynalski, K., Rossi, P., and Holliger, C.** (2011) Reductive dechlorination of tetrachloroethene by a stepwise catalysis of different organohalide respiring bacteria and reductive dehalogenases. *Biodegradation*: doi: 10.1007/s10532-10011-19454-10534.
- Major, D.W., McMaster, M.L., Cox, E.E., Edwards, E.A., Dworatzek, S.M., Hendrickson, E.R. et al.** (2002) Field Demonstration of Successful Bioaugmentation To Achieve Dechlorination of Tetrachloroethene To Ethene. *Environ Sci Technol* **36**: 5106-5116.
- Maphosa, F., de Vos, W.M., and Smidt, H.** (2010) Exploiting the ecogenomics toolbox for environmental diagnostics of organohalide-respiring bacteria. *Trends Biotechnol* **28**: 308-316.
- Mars, A.E., Prins, G.T., Wietzes, P., de Koning, W., and Janssen, D.B.** (1998) Effect of trichloroethylene on the competitive behavior of toluene-degrading bacteria. *Appl Environ Microbiol* **64**: 208-215.

- Mattes, T.E., Alexander, A.K., and Coleman, N.V.** (2010) Aerobic biodegradation of the chloroethenes: pathways, enzymes, ecology, and evolution. *FEMS Microbiol Rev* **34**: 445-475.
- Mattes, T.E., Coleman, N.V., Spain, J.C., and Gossett, J.M.** (2005) Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in *Nocardioides* sp. strain JS614. *Arch Microbiol* **183**: 95-106.
- Mattes, T.E., Alexander, A.K., Richardson, P.M., Munk, A.C., Han, C.S., Stothard, P., and Coleman, N.V.** (2008) The genome of *Polaromonas* sp. strain JS666: insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. *Appl Environ Microbiol* **74**: 6405-6416.
- Maupin-Furlow, J.A., and Ferry, J.G.** (1996) Analysis of the CO dehydrogenase/acetyl-coenzyme A synthase operon of *Methanosarcina thermophila*. *J Bacteriol* **178**: 6849-6856.
- May, H.D., Miller, G.S., Kjellerup, B.V., and Sowers, K.R.** (2008) Dehalorespiration with polychlorinated biphenyls by an anaerobic ultramicrobacterium. *Appl Environ Microbiol* **74**: 2089-2094.
- May, H.D., Cutter, L.A., Miller, G.S., Milliken, C.E., Watts, J.E., and Sowers, K.R.** (2006) Stimulatory and inhibitory effects of organohalides on the dehalogenating activities of PCB-dechlorinating bacterium o-17. *Environ Sci Technol* **40**: 5704-5709.
- Maymo-Gatell, X., Nijenhuis, I., and Zinder, S.H.** (2001) Reductive dechlorination of cis-1,2-dichloroethene and vinyl chloride by "Dehalococcoides ethenogenes". *Environ Sci Technol* **35**: 516-521.
- Maymo-Gatell, X., Tandoi, V., Gossett, J.M., and Zinder, S.H.** (1995) Characterization of an H₂-utilizing enrichment culture that reductively dechlorinates tetrachloroethene to vinyl chloride and ethene in the absence of methanogenesis and acetogenesis. *Appl Environ Microbiol* **61**: 3928-3933.
- Maymo-Gatell, X., Chien, Y., Gossett, J.M., and Zinder, S.H.** (1997) Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**: 1568-1571.
- McCarthy, A.** (2010) Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol* **17**: 675-676.
- McCarty, P.L.** (1997) Breathing with chlorinated solvents. *Science* **276**: 1521-1522.
- McMurdie, P.J., Behrens, S.F., Holmes, S., and Spormann, A.M.** (2007) Unusual codon bias in vinyl chloride reductase genes of *Dehalococcoides* species. *Appl Environ Microbiol* **73**: 2744-2747.
- McMurdie, P.J., Behrens, S.F., Muller, J.A., Goke, J., Ritalahti, K.M., Wagner, R. et al.** (2009) Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genet* **5**: e1000714.
- Meckenstock, R.U., Morasch, B., Griebler, C., and Richnow, H.H.** (2004) Stable isotope fractionation analysis as a tool to monitor biodegradation in contaminated aquifers. *J Contam Hydrol* **75**: 215-255.
- Megharaj, M., Ramakrishnan, B., Venkateswarlu, K., Sethunathan, N., and Naidu, R.** (2011) Bioremediation approaches for organic pollutants: A critical perspective. *Environ Int* **37**: 1362-1375.
- Metzker, M.L.** (2010) Sequencing technologies - the next generation. *Nat Rev Genet* **11**: 31-46.
- Milaton, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D. et al.** (2007) PhylArray: phylogenetic probe design algorithm for microarray. *Bioinformatics* **23**: 2550-2557.
- Miller, E., Wohlfarth, G., and Diekert, G.** (1998) Purification and characterization of the tetrachloroethene reductive dehalogenase of strain PCE-S. *Arch Microbiol* **169**: 497-502.

- Miller, G.S., Milliken, C.E., Sowers, K.R., and May, H.D.** (2005) Reductive dechlorination of tetrachloroethene to trans-dichloroethene and cis-dichloroethene by PCB-dechlorinating bacterium DF-1. *Environ Sci Technol* **39**: 2631-2635.
- Miyata, R., Adachi, K., Tani, H., Kurata, S., Nakamura, K., Tsuneda, S. et al.** (2010) Quantitative detection of chloroethene-reductive bacteria *Dehalococcoides* spp. using alternately binding probe competitive Polymerase Chain Reaction. *Mol Cell Probes* **24**: 131-137.
- Moe, W.M., Yan, J., Nobre, M.F., da Costa, M.S., and Rainey, F.A.** (2009) *Dehalogenimonas lykanthroporepellens* gen. nov., sp. nov., a reductively dehalogenating bacterium isolated from chlorinated solvent-contaminated groundwater. *Int J Syst Evol Microbiol* **59**: 2692-2697.
- Morales, S.E., and Holben, W.E.** (2010) Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? *FEMS Microbiol Ecol* **75**: 2-16.
- Mravik, S.C., Sillan, R.K., Wood, A.L., and Sewell, G.W.** (2003) Field evaluation of the solvent extraction residual biotreatment technology. *Environ Sci Technol* **37**: 5040-5049.
- Mueckstein, U., Lepar, G.G., Posekany, A., Hofacker, I., and Kreil, D.P.** (2010) Hybridization thermodynamics of NimbleGen microarrays. *BMC Bioinformatics* **11**: 35.
- Muller, J.A., Rosner, B.M., Von Abendroth, G., Meshulam-Simon, G., McCarty, P.L., and Spormann, A.M.** (2004) Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. strain VS and its environmental distribution. *Appl Environ Microbiol* **70**: 4880-4888.
- Munroe, D.J., and Harris, T.J.** (2010) Third-generation sequencing fireworks at Marco Island. *Nat Biotechnol* **28**: 426-428.
- Nemir, A., David, M.M., Perrussel, R., Sapkota, A., Simonet, P., Monier, J.M., and Vogel, T.M.** (2010) Comparative phylogenetic microarray analysis of microbial communities in TCE-contaminated soils. *Chemosphere* **80**: 600-607.
- Neufeld, J.D., Mohn, W.W., and de Lorenzo, V.** (2006) Composition of microbial communities in hexachlorocyclohexane (HCH) contaminated soils from Spain revealed with a habitat-specific microarray. *Environ Microbiol* **8**: 126-140.
- Neumann, A., Wohlfarth, G., and Diekert, G.** (1996) Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*. *J Biol Chem* **271**: 16515-16519.
- Nevin, K.P., Holmes, D.E., Woodard, T.L., Covalla, S.F., and Lovley, D.R.** (2007) Reclassification of *Trichlorobacter thiogenes* as *Geobacter thiogenes* comb. nov. *Int J Syst Evol Microbiol* **57**: 463-466.
- Newman, L.A., Strand, S.E., Choe, N., Duffy, J., Ekuan, G., Ruszaj, M. et al.** (1997) Uptake and Biotransformation of Trichloroethylene by Hybrid Poplars. *Environ Sci Technol* **31**: 1062-1067.
- Nicolaou, K.C., Boddy, C.N., Brase, S., and Winssinger, N.** (1999) Chemistry, Biology, and Medicine of the Glycopeptide Antibiotics. *Angew Chem Int Ed Engl* **38**: 2096-2152.
- Nietz, C.T., Morris, J.T., and Vroblesky, D.A.** (1999) Biophysical Mechanisms of Trichloroethene Uptake and Loss in Baldcypress Growing in Shallow Contaminated Groundwater. *Environ Sci Technol* **33**: 2899-2904.
- Nishimura, M., Ebisawa, M., Sakihara, S., Kobayashi, A., Nakama, T., Okochi, M., and Yohda, M.** (2008) Detection and identification of *Dehalococcoides* species responsible for in situ dechlorination of trichloroethene to ethene enhanced by hydrogen-releasing compounds. *Biotechnol Appl Biochem* **51**: 1-7.
- Nobre, R.C., and Nobre, M.M.** (2004) Natural attenuation of chlorinated organics in a shallow sand aquifer. *J Hazard Mater* **110**: 129-137.

- Noireaux, V., Maeda, Y.T., and Libchaber, A.** (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc Natl Acad Sci U S A* **108**: 3473-3480.
- Oberg, G.** (2002) The natural chlorine cycle--fitting the scattered pieces. *Appl Microbiol Biotechnol* **58**: 565-581.
- Oldenhuis, R., Vink, R.L., Janssen, D.B., and Witholt, B.** (1989) Degradation of chlorinated aliphatic hydrocarbons by *Methylosinus trichosporium* OB3b expressing soluble methane monooxygenase. *Appl Environ Microbiol* **55**: 2819-2826.
- Pace, N.R.** (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**: 734-740.
- Pant, P., and Pant, S.** (2010) A review: advances in microbial remediation of trichloroethylene (TCE). *J Environ Sci (China)* **22**: 116-126.
- Pareek, C.S., Smoczynski, R., and Tretyn, A.** (2011) Sequencing technologies and genome sequencing. *J Appl Genet.*
- Pathak, A., Shanker, R., Garg, S.K., and Manickam, N.** (2011) Profiling of biodegradation and bacterial 16S rRNA genes in diverse contaminated ecosystems using 60-mer oligonucleotide microarray. *Appl Microbiol Biotechnol* **90**: 1739-1754.
- Pemont, J.** (2005) *Biodegradation et métabolismes*.
- Perelo, L.W.** (2010) Review: In situ and bioremediation of organic pollutants in aquatic sediments. *J Hazard Mater* **177**: 81-89.
- Petkau, A., Stuart-Edwards, M., Stothard, P., and Van Domselaar, G.** (2010) Interactive microbial genome visualization with GView. *Bioinformatics* **26**: 3125-3126.
- Pierson, B.K., and Castenholz, R.W.** (1974) A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Arch Microbiol* **100**: 5-24.
- Pikuta, E.V., Hoover, R.B., and Tang, J.** (2007) Microbial extremophiles at the limits of life. *Crit Rev Microbiol* **33**: 183-209.
- Pilon-Smits, E.** (2005) Phytoremediation. *Annu Rev Plant Biol* **56**: 15-39.
- Pinedo-Rivilla, C., Aleu, J., and Collado, I.G.** (2009) Pollutants biodegradation by fungi. *Curr Org Chem* **13**: 1194-1214.
- Polz, M.F., and Cavanaugh, C.M.** (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724-3730.
- Pozhitkov, A.E., Tautz, D., and Noble, P.A.** (2007) Oligonucleotide microarrays: widely applied poorly understood. *Brief Funct Genomic Proteomic* **6**: 141-148.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O.** (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188-7196.
- Quince, C., Curtis, T.P., and Sloan, W.T.** (2008) The rational exploration of microbial diversity. *Isme J* **2**: 997-1006.
- Quinn, J., Geiger, C., Clausen, C., Brooks, K., Coon, C., O'Hara, S. et al.** (2005) Field Demonstration of DNAPL Dehalogenation Using Emulsified Zero-Valent Iron. *Environ Sci Technol* **39**: 1309-1318.
- Rahm, B.G., Morris, R.M., and Richardson, R.E.** (2006) Temporal expression of respiratory genes in an enrichment culture containing *Dehalococcoides ethenogenes*. *Appl Environ Microbiol* **72**: 5486-5491.
- Ramos, J.L., Marques, S., van Dillewijn, P., Espinosa-Urgel, M., Segura, A., Duque, E. et al.** (2011) Laboratory research aimed at closing the gaps in microbial bioremediation. *Trends Biotechnol.*
- Ramsburg, C.A., Abriola, L.M., Pennell, K.D., Löffler, F.E., Gamache, M., Amos, B.K., and Petrovskis, E.A.** (2004) Stimulated Microbial Reductive Dechlorination following Surfactant Treatment at the Bachman Road Site. *Environ Sci Technol* **38**: 5902-5914.

- Rastogi, G., Barua, S., Sani, R.K., and Peyton, B.M.** (2011) Investigation of Microbial Populations in the Extremely Metal-Contaminated Coeur d'Alene River Sediments. *Microb Ecol* **62**: 1-13.
- Rastogi, G., Osman, S., Vaishampayan, P.A., Andersen, G.L., Stetler, L.D., and Sani, R.K.** (2010a) Microbial diversity in uranium mining-impacted soils as revealed by high-density 16S microarray and clone library. *Microb Ecol* **59**: 94-108.
- Rastogi, G., Osman, S., Kukkadapu, R., Engelhard, M., Vaishampayan, P.A., Andersen, G.L., and Sani, R.K.** (2010b) Microbial and mineralogical characterizations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microb Ecol* **60**: 539-550.
- Regeard, C., Maillard, J., and Holliger, C.** (2004) Development of degenerate and specific PCR primers for the detection and isolation of known and putative chloroethene reductive dehalogenase genes. *J Microbiol Methods* **56**: 107-118.
- Regeard, C., Maillard, J., Dufraigne, C., Deschavanne, P., and Holliger, C.** (2005) Indications for acquisition of reductive dehalogenase genes through horizontal gene transfer by Dehalococcoides ethenogenes strain 195. *Appl Environ Microbiol* **71**: 2955-2961.
- Religio, A., Schwager, C., Richter, A., Ansorge, W., and Valcarcel, J.** (2002) Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res* **30**: e51.
- Rhee, S.-K., Liu, X., Wu, L., Chong, S.C., Wan, X., and Zhou, J.** (2004) Detection of Genes Involved in Biodegradation and Biotransformation in Microbial Communities by Using 50-Mer Oligonucleotide Microarrays. *Appl Environ Microbiol* **70**: 4303-4317.
- Richardson, R.E., James, C.A., Bhupathiraju, V.K., and Alvarez-Cohen, L.** (2002) Microbial activity in soils following steam treatment. *Biodegradation* **13**: 285-295.
- Rieger, P.G., Meier, H.M., Gerle, M., Vogt, U., Groth, T., and Knackmuss, H.J.** (2002) Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. *J Biotechnol* **94**: 101-123.
- Rinta-Kanto, J.M., Burgmann, H., Gifford, S.M., Sun, S., Sharma, S., del Valle, D.A. et al.** (2011) Analysis of sulfur-related transcription by Roseobacter communities using a taxon-specific functional gene microarray. *Environ Microbiol* **13**: 453-467.
- Ritalahti, K.M., Amos, B.K., Sung, Y., Wu, Q., Koenigsberg, S.S., and Löffler, F.E.** (2006) Quantitative PCR Targeting 16S rRNA and Reductive Dehalogenase Genes Simultaneously Monitors Multiple Dehalococcoides Strains. *Appl Environ Microbiol* **72**: 2765-2774.
- Roberts, A.L., Totten, L.A., Arnold, W.A., Burris, D.R., and Campbell, T.J.** (1996) Reductive Elimination of Chlorinated Ethylenes by Zero-Valent Metals. *Environ Sci Technol* **30**: 2654-2659.
- Robinson, J.P., Kingman, S.W., and Onobrakpeya, O.** (2008) Microwave-assisted stripping of oil contaminated drill cuttings. *J Environ Manage* **88**: 211-218.
- Rocha, E.P., and Danchin, A.** (2003a) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* **31**: 6570-6577.
- Rocha, E.P., and Danchin, A.** (2003b) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* **34**: 377-378.
- Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R. et al.** (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**: 2541-2547.
- Rose, T.M., Henikoff, J.G., and Henikoff, S.** (2003) CODEHOP (Consensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* **31**: 3763-3766.
- Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S.** (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**: 1628-1635.

- Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M. et al.** (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**: 348-352.
- Rouzeau-Szynalski, K., Maillard, J., and Holliger, C.** (2011) Frequent concomitant presence of *Desulfitobacterium* spp. and "Dehalococcoides" spp. in chloroethene-dechlorinating microbial communities. *Appl Microbiol Biotechnol* **90**: 361-368.
- Ruder, A.M.** (2006) Potential health effects of occupational chlorinated solvent exposure. *Ann N Y Acad Sci* **1076**: 207-227.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S. et al.** (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: e77.
- Saeki, H., Akira, M., Furuhashi, K., Averhoff, B., and Gottschalk, G.** (1999) Degradation of trichloroethene by a linear-plasmid-encoded alkene monooxygenase in *Rhodococcus corallinus* (*Nocardia corallina*) B-276. *Microbiology* **145** (Pt 7): 1721-1730.
- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N.** (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354.
- Sanguin, H., Sarniguet, A., Gazengel, K., Moenne-Loccoz, Y., and Grundmann, G.L.** (2009) Rhizosphere bacterial communities associated with disease suppressiveness stages of take-all decline in wheat monoculture. *New Phytol* **184**: 694-707.
- Sanguin, H., Remenant, B., Dechesne, A., Thioulouse, J., Vogel, T.M., Nesme, X. et al.** (2006a) Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. *Appl Environ Microbiol* **72**: 4302-4312.
- Sanguin, H., Herrera, A., Oger-Desfeux, C., Dechesne, A., Simonet, P., Navarro, E. et al.** (2006b) Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria. *Environ Microbiol* **8**: 289-307.
- Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., and Nolan, G.P.** (2010) Computational solutions to large-scale data management and analysis. *Nat Rev Genet* **11**: 647-657.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O.** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470.
- Scheutz, C., Durant, N.D., Dennis, P., Hansen, M.H., Jorgensen, T., Jakobsen, R. et al.** (2008) Concurrent ethene generation and growth of *Dehalococcoides* containing vinyl chloride reductive dehalogenase genes during an enhanced reductive dechlorination field demonstration. *Environ Sci Technol* **42**: 9302-9309.
- Scheutz, C., Broholm, M.M., Durant, N.D., Weeth, E.B., Jorgensen, T.H., Dennis, P. et al.** (2010) Field evaluation of biological enhanced reductive dechlorination of chloroethenes in clayey till. *Environ Sci Technol* **44**: 5134-5141.
- Schnabel, W.E., Dietz, A.C., Burken, J.G., Schnoor, J.L., and Alvarez, P.J.** (1997) Uptake and transformation of trichloroethylene by edible garden plants. *Water Res* **31**: 816-824.
- Scholz-Muramatsu, H., Neumann, A., Meßmer, M., Moore, E., and Diekert, G.** (1995) Isolation and characterization of *Dehalospirillum multivorans* gen. nov., sp. nov., a tetrachloroethene-utilizing, strictly anaerobic bacterium. *Arch Microbiol* **163**: 48-56.
- Schonhuber, W., Fuchs, B., Juretschko, S., and Amann, R.** (1997) Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. *Appl Environ Microbiol* **63**: 3268-3273.

- Schonmann, S., Loy, A., Wimmersberger, C., Sobek, J., Aquino, C., Vandamme, P. et al.** (2009) 16S rRNA gene-based phylogenetic microarray for simultaneous identification of members of the genus Burkholderia. *Environ Microbiol* **11**: 779-800.
- Schroth, M.H., Oostrom, M., Wietsma, T.W., and Istok, J.D.** (2001) In-situ oxidation of trichloroethene by permanganate: effects on porous medium hydraulic properties. *J Contam Hydrol* **50**: 79-98.
- Schumacher, W., and Holliger, C.** (1996) The proton/electron ration of the menaquinone-dependent electron transport from dihydrogen to tetrachloroethene in "Dehalobacter restrictus". *J Bacteriol* **178**: 2328-2333.
- Scullion, J.** (2006) Remediating polluted soils. *Naturwissenschaften* **93**: 51-65.
- Semprini, L.** (1995) In situ bioremediation of chlorinated solvents. *Environ Health Perspect* **103 Suppl 5**: 101-105.
- Seshadri, R., Adrian, L., Fouts, D.E., Eisen, J.A., Phillippy, A.M., Methe, B.A. et al.** (2005) Genome sequence of the PCE-dechlorinating bacterium Dehalococcoides ethenogenes. *Science* **307**: 105-108.
- Shim, H., Ryoo, D., Barbieri, P., and Wood, T.K.** (2001) Aerobic degradation of mixtures of tetrachloroethylene, trichloroethylene, dichloroethylenes, and vinyl chloride by toluene-o-xylene monooxygenase of Pseudomonas stutzeri OX1. *Appl Microbiol Biotechnol* **56**: 265-269.
- Smets, B.F., and Pritchard, P.H.** (2003) Elucidating the microbial component of natural attenuation. *Curr Opin Biotechnol* **14**: 283-288.
- Smidt, H., and de Vos, W.M.** (2004) Anaerobic microbial dehalogenation. *Annu Rev Microbiol* **58**: 43-73.
- Smits, T.H., Devenoges, C., Szynalski, K., Maillard, J., and Holliger, C.** (2004) Development of a real-time PCR method for quantification of the three genera Dehalobacter, Dehalococcoides, and Desulfitobacterium in microbial communities. *J Microbiol Methods* **57**: 369-378.
- Spande, T.F., Garraffo, H.M., Edwards, M.W., Yeh, H.J.C., Pannell, L., and Daly, J.W.** (1992) Epibatidine: a novel (chloropyridyl)azabicycloheptane with potent analgesic activity from an Ecuadorian poison frog. *J Am Chem Soc* **114**: 3475-3478.
- Stralis-Pavese, N., Abell, G.C., Sessitsch, A., and Bodrossy, L.** (2011) Analysis of methanotroph community composition using a pmoA-based microbial diagnostic microarray. *Nat Protoc* **6**: 609-624.
- Sung, Y., Ritalahti, K.M., Apkarian, R.P., and Löffler, F.E.** (2006a) Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring Dehalococcoides isolate. *Appl Environ Microbiol* **72**: 1980-1987.
- Sung, Y., Ritalahti, K.M., Sanford, R.A., Urbance, J.W., Flynn, S.J., Tiedje, J.M., and Löffler, F.E.** (2003) Characterization of two tetrachloroethene-reducing, acetate-oxidizing anaerobic bacteria and their description as Desulfuromonas michiganensis sp. nov. *Appl Environ Microbiol* **69**: 2964-2974.
- Sung, Y., Fletcher, K.E., Ritalahti, K.M., Apkarian, R.P., Ramos-Hernandez, N., Sanford, R.A. et al.** (2006b) Geobacter lovleyi sp. nov. strain SZ, a novel metal-reducing and tetrachloroethene-dechlorinating bacterium. *Appl Environ Microbiol* **72**: 2775-2782.
- Suyama, A., Yamashita, M., Yoshino, S., and Furukawa, K.** (2002) Molecular characterization of the PceA reductive dehalogenase of desulfitobacterium sp. strain Y51. *J Bacteriol* **184**: 3419-3425.
- Suyama, A., Iwakiri, R., Kai, K., Tokunaga, T., Sera, N., and Furukawa, K.** (2001) Isolation and characterization of Desulfitobacterium sp. strain Y51 capable of efficient dehalogenation of tetrachloroethene and polychloroethanes. *Biosci Biotechnol Biochem* **65**: 1474-1481.

Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625-630.

Takeuchi, M., Kawabe, Y., Watanabe, E., Oiwa, T., Takahashi, M., Nanba, K. et al. (2011) Comparative study of microbial dechlorination of chlorinated ethenes in an aquifer and a clayey aquitard. *J Contam Hydrol* **124**: 14-24.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.

Tao, W., Lee, M.H., Wu, J., Kim, N.H., and Lee, S.W. (2011) Isolation and characterization of a family VII esterase derived from alluvial soil metagenomic library. *J Microbiol* **49**: 178-185.

Taroncher-Oldenburg, G., Griner, E.M., Francis, C.A., and Ward, B.B. (2003) Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* **69**: 1159-1171.

Tas, N., van Eekert, M.H., de Vos, W.M., and Smidt, H. (2010) The little bacteria that can - diversity, genomics and ecophysiology of 'Dehalococcoides' spp. in contaminated environments. *Microb Biotechnol* **3**: 389-402.

Tas, N., van Eekert, M.H., Schraa, G., Zhou, J., de Vos, W.M., and Smidt, H. (2009) Tracking functional guilds: "Dehalococcoides" spp. in European river basins contaminated with hexachlorobenzene. *Appl Environ Microbiol* **75**: 4696-4704.

Terrat, S., Peyretailade, E., Goncalves, O., Dugat-Bony, E., Gravelat, F., Mone, A. et al. (2010) Detecting variants with Metabolic Design, a new software tool to design probes for explorative functional DNA microarray development. *BMC Bioinformatics* **11**: 478.

Tiehm, A., and Schmidt, K.R. (2011) Sequential anaerobic/aerobic biodegradation of chloroethenes--aspects of field application. *Curr Opin Biotechnol* **22**: 415-421.

Timp, W., Mirsaidov, U.M., Wang, D., Comer, J., Aksimentiev, A., and Timp, G. (2010) Nanopore Sequencing: Electrical Measurements of the Code of Life. *IEEE Trans Nanotechnol* **9**: 281-294.

Tiquia, S.M., Wu, L., Chong, S.C., Passovets, S., Xu, D., Xu, Y., and Zhou, J. (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **36**: 664-670, 672, 674-665.

Torsvik, V., and Øvreås, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240-245.

Torsvik, V., Goksoyr, J., and Daae, F.L. (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56**: 782-787.

Travis, C., and Doty, C. (1990) ES&T Views: Can contaminated aquifers at superfund sites be remediated? *Environ Sci Technol* **24**: 1464-1466.

Truex, M.J., Macbeth, T.W., Vermeul, V.R., Fritz, B.G., Mendoza, D.P., Mackley, R.D. et al. (2011) Demonstration of combined zero-valent iron and electrical resistance heating for in situ trichloroethene remediation. *Environ Sci Technol* **45**: 5346-5351.

Tsai, T.T., Kao, C.M., Yeh, T.Y., Liang, S.H., and Chien, H.Y. (2009) Application of surfactant enhanced permanganate oxidation and biodegradation of trichloroethylene in groundwater. *J Hazard Mater* **161**: 111-119.

Tsien, H.C., Brusseau, G.A., Hanson, R.S., and Waclett, L.P. (1989) Biodegradation of trichloroethylene by *Methylosinus trichosporium* OB3b. *Appl Environ Microbiol* **55**: 3155-3161.

Tsukagoshi, N., Ezaki, S., Uenaka, T., Suzuki, N., and Kurane, R. (2006) Isolation and transcriptional analysis of novel tetrachloroethene reductive dehalogenase gene from *Desulfitobacterium* sp. strain KBC1. *Appl Microbiol Biotechnol* **69**: 543-553.

- Tyagi, M., da Fonseca, M.M., and de Carvalho, C.C.** (2011) Bioaugmentation and biostimulation strategies to improve the effectiveness of bioremediation processes. *Biodegradation* **22**: 231-241.
- Vainberg, S., Condee, C., and Steffan, R.** (2009) Large-scale production of bacterial consortia for remediation of chlorinated solvent-contaminated groundwater. *J Ind Microbiol Biotechnol* **36**: 1189-1197.
- Valm, A.M., Welch, J.L., Rieken, C.W., Hasegawa, Y., Sogin, M.L., Oldenbourg, R. et al.** (2011) Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc Natl Acad Sci U S A* **108**: 4152-4157.
- van de Pas, B.A., Gerritse, J., de Vos, W.M., Schraa, G., and Stams, A.J.** (2001) Two distinct enzyme systems are responsible for tetrachloroethene and chlorophenol reductive dehalogenation in *Desulfitobacterium* strain PCE1. *Arch Microbiol* **176**: 165-169.
- van der Meer, J.R., de Vos, W.M., Harayama, S., and Zehnder, A.J.** (1992) Molecular mechanisms of genetic adaptation to xenobiotic compounds. *Microbiol Rev* **56**: 677-694.
- van der Zaan, B., Hannes, F., Hoekstra, N., Rijnaarts, H., de Vos, W.M., Smidt, H., and Gerritse, J.** (2010) Correlation of Dehalococcoides 16S rRNA and Chloroethene-Reductive Dehalogenase Genes with Geochemical Conditions in Chloroethene-Contaminated Groundwater. *Appl. Environ. Microbiol.* **76**: 843-850.
- van Hylckama Vlieg, J.E., and Janssen, D.B.** (2001) Formation and detoxification of reactive intermediates in the metabolism of chlorinated ethenes. *J Biotechnol* **85**: 81-102.
- Van Nostrand, J.D., Wu, W.-M., Wu, L., Deng, Y., Carley, J., Carroll, S. et al.** (2009) GeoChip-based analysis of functional microbial communities during the reoxidation of a bioreduced uranium-contaminated aquifer. *Environ Microbiol* **11**: 2611-2626.
- Van Nostrand, J.D., Wu, L., Wu, W.M., Huang, Z., Gentry, T.J., Deng, Y. et al.** (2011) Dynamics of microbial community composition and function during in situ bioremediation of a uranium-contaminated aquifer. *Appl Environ Microbiol* **77**: 3860-3869.
- Vandecasteele, J.** (2005) *Microbiologie Pétrolière*: IFP Publications.
- Vartoukian, S.R., Palmer, R.M., and Wade, W.G.** (2010) Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol Lett* **309**: 1-7.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. et al.** (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.
- Verce, M.F., Ulrich, R.L., and Freedman, D.L.** (2000) Characterization of an isolate that uses vinyl chloride as a growth substrate under aerobic conditions. *Appl Environ Microbiol* **66**: 3535-3542.
- Verce, M.F., Ulrich, R.L., and Freedman, D.L.** (2001) Transition from cometabolic to growth-linked biodegradation of vinyl chloride by a *Pseudomonas* sp. isolated on ethene. *Environ Sci Technol* **35**: 4242-4251.
- Vidali, M.** (2001) Bioremediation. An overview. *Pure Appl Chem* **73**: 1163-1172.
- Vignais, P.M., and Billoud, B.** (2007) Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**: 4206-4272.
- Villemur, R., Lanthier, M., Beaudet, R., and Lepine, F.** (2006) The *Desulfitobacterium* genus. *FEMS Microbiol Rev* **30**: 706-733.
- Vogel, T.M.** (1996) Bioaugmentation as a soil bioremediation approach. *Curr Opin Biotechnol* **7**: 311-316.
- Vogel, T.M., and McCarty, P.L.** (1985) Biotransformation of tetrachloroethylene to trichloroethylene, dichloroethylene, vinyl chloride, and carbon dioxide under methanogenic conditions. *Appl Environ Microbiol* **49**: 1080-1083.
- Vogel, T.M., Criddle, C.S., and McCarty, P.L.** (1987) ES Critical Reviews: Transformations of halogenated aliphatic compounds. *Environ Sci Technol* **21**: 722-736.

- Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M., van Elsas, J.D. et al.** (2009) TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat Rev Micro* **7**: 252-252.
- von Wintzingerode, F., Gobel, U.B., and Stackebrandt, E.** (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213-229.
- von Wintzingerode, F., Schlotelburg, C., Hauck, R., Hegemann, W., and Gobel, U.B.** (2001) Development of primers for amplifying genes encoding CprA- and PceA-like reductive dehalogenases in anaerobic microbial consortia, dechlorinating trichlorobenzene and 1,2-dichloropropane. *FEMS Microbiol Ecol* **35**: 189-196.
- Vroblesky, D.A., Nietch, C.T., and Morris, J.T.** (1998) Chlorinated Ethenes from Groundwater in Tree Trunks. *Environ Sci Technol* **33**: 510-515.
- Wackett, L.P., and Gibson, D.T.** (1988) Degradation of trichloroethylene by toluene dioxygenase in whole-cell studies with *Pseudomonas putida* F1. *Appl Environ Microbiol* **54**: 1703-1708.
- Wackett, L.P., Brusseau, G.A., Householder, S.R., and Hanson, R.S.** (1989) Survey of microbial oxygenases: trichloroethylene degradation by propane-oxidizing bacteria. *Appl Environ Microbiol* **55**: 2960-2964.
- Wagner, M., Smidt, H., Loy, A., and Zhou, J.** (2007) Unravelling Microbial Communities with DNA-Microarrays: Challenges and Future Directions. *Microb Ecol* **53**: 498-506.
- Ward, B.B., Eveillard, D., Kirshtein, J.D., Nelson, J.D., Voytek, M.A., and Jackson, G.A.** (2007) Ammonia-oxidizing bacterial community composition in estuarine and oceanic environments assessed using a functional gene microarray. *Environ Microbiol* **9**: 2522-2538.
- Watts, J.E., Fagervold, S.K., May, H.D., and Sowers, K.R.** (2005) A PCR-based specific assay reveals a population of bacteria within the Chloroflexi associated with the reductive dehalogenation of polychlorinated biphenyls. *Microbiology* **151**: 2039-2046.
- Whitman, W.B., Coleman, D.C., and Wiebe, W.J.** (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci U S A* **95**: 6578-6583.
- Wilmes, P., and Bond, P.L.** (2004) The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* **6**: 911-920.
- Wilson, J.T., and Wilson, B.H.** (1985) Biotransformation of trichloroethylene in soil. *Appl Environ Microbiol* **49**: 242-243.
- Winter, R.B., Yen, K.-M., and Ensley, B.D.** (1989) Efficient Degradation of Trichloroethylene by a Recombinant *Escherichia Coli*. *Nat Biotechnol* **7**: 282-285.
- Woese, C.R., Kandler, O., and Wheelis, M.L.** (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.
- Wu, L., Thompson, D.K., Li, G., Hurt, R.A., Tiedje, J.M., and Zhou, J.** (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* **67**: 5780-5790.
- Wu, Q., Sowers, K.R., and May, H.D.** (2000) Establishment of a polychlorinated biphenyl-dechlorinating microbial consortium, specific for doubly flanked chlorines, in a defined, sediment-free medium. *Appl Environ Microbiol* **66**: 49-53.
- Wu, Q., Watts, J.E., Sowers, K.R., and May, H.D.** (2002a) Identification of a bacterium that specifically catalyzes the reductive dechlorination of polychlorinated biphenyls with doubly flanked chlorines. *Appl Environ Microbiol* **68**: 807-812.
- Wu, Q., Milliken, C.E., Meier, G.P., Watts, J.E., Sowers, K.R., and May, H.D.** (2002b) Dechlorination of chlorobenzenes by a culture containing bacterium DF-1, a PCB dechlorinating microorganism. *Environ Sci Technol* **36**: 3290-3294.

- Xie, J., He, Z., Liu, X., Liu, X., Van Nostrand, J.D., Deng, Y. et al.** (2011) GeoChip-based analysis of the functional gene diversity and metabolic potential of microbial communities in acid mine drainage. *Appl Environ Microbiol* **77**: 991-999.
- Xiong, J., Wu, L., Tu, S., Van Nostrand, J.D., He, Z., Zhou, J., and Wang, G.** (2010) Microbial communities and functional genes associated with soil arsenic contamination and the rhizosphere of the arsenic-hyperaccumulating plant *Pteris vittata* L. *Appl Environ Microbiol* **76**: 7277-7284.
- Xu, J.** (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* **15**: 1713-1731.
- Xu, M., Wu, W.M., Wu, L., He, Z., Van Nostrand, J.D., Deng, Y. et al.** (2010) Responses of microbial community functional structures to pilot-scale uranium in situ bioremediation. *Isme J* **4**: 1060-1070.
- Yabe, S., Aiba, Y., Sakai, Y., Hazaka, M., and Yokota, A.** (2010) *Thermosporothrix hazakensis* gen. nov., sp. nov., isolated from compost, description of *Thermosporotrichaceae* fam. nov. within the class *Ktedonobacteria* Cavaletti et al. 2007 and emended description of the class *Ktedonobacteria*. *Int J Syst Evol Microbiol* **60**: 1794-1801.
- Yamada, T., and Sekiguchi, Y.** (2009) Cultivation of uncultured chloroflexi subphyla: significance and ecophysiology of formerly uncultured chloroflexi 'subphylum i' with natural and biotechnological relevance. *Microbes Environ* **24**: 205-216.
- Yan, J., Rash, B.A., Rainey, F.A., and Moe, W.M.** (2009a) Isolation of novel bacteria within the Chloroflexi capable of reductive dechlorination of 1,2,3-trichloropropane. *Environ Microbiol* **11**: 833-843.
- Yan, J., Rash, B.A., Rainey, F.A., and Moe, W.M.** (2009b) Detection and quantification of Dehalogenimonas and "Dehalococcoides" populations via PCR-based protocols targeting 16S rRNA genes. *Appl Environ Microbiol* **75**: 7560-7564.
- Yan, T., LaPara, T.M., and Novak, P.J.** (2006) The reductive dechlorination of 2,3,4,5-tetrachlorobiphenyl in three different sediment cultures: evidence for the involvement of phylogenetically similar Dehalococcoides-like bacterial populations. *FEMS Microbiol Ecol* **55**: 248-261.
- Yang, Y., and McCarty, P.L.** (1998) Competition for Hydrogen within a Chlorinated Solvent Dehalogenating Anaerobic Mixed Culture. *Environ Sci Technol* **32**: 3591-3597.
- Yang, Y., and Zeyer, J.** (2003) Specific detection of Dehalococcoides species by fluorescence in situ hybridization with 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* **69**: 2879-2883.
- Yang, Y., Pesaro, M., Sigler, W., and Zeyer, J.** (2005) Identification of microorganisms involved in reductive dehalogenation of chlorinated ethenes in an anaerobic microbial community. *Water Res* **39**: 3954-3966.
- Yee, D.C., Maynard, J.A., and Wood, T.K.** (1998) Rhizoremediation of trichloroethylene by a recombinant, root-colonizing *Pseudomonas fluorescens* strain expressing toluene ortho-monooxygenase constitutively. *Appl Environ Microbiol* **64**: 112-118.
- Yoshida, N., Takahashi, N., and Hiraishi, A.** (2005) Phylogenetic characterization of a polychlorinated-dioxin- dechlorinating microbial community by use of microcosm studies. *Appl Environ Microbiol* **71**: 4325-4334.
- Yu, S., and Semprini, L.** (2009) Enhanced reductive dechlorination of PCE DNAPL with TBOS as a slow-release electron donor. *J Hazard Mater* **167**: 97-104.
- Zhao, B., Zhu, L., and Yang, K.** (2006) Solubilization of DNAPLs by mixed surfactant: reduction in partitioning losses of nonionic surfactant. *Chemosphere* **62**: 772-779.
- Zhou, J.** (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288-294.

Ziv-El, M., Delgado, A.G., Yao, Y., Kang, D.W., Nelson, K.G., Halden, R.U., and Krajmalnik-Brown, R. (2011) Development and characterization of DehaloR², a novel anaerobic microbial consortium performing rapid dechlorination of TCE to ethene. *Appl Microbiol Biotechnol*.

ANNEXES

RESEARCH ARTICLE

Open Access

Detecting variants with Metabolic Design, a new software tool to design probes for explorative functional DNA microarray development

Sébastien Terrat^{1,2,3}, Eric Peyretailade^{1,2}, Olivier Gonçalves^{1,2}, Eric Dugat-Bony^{2,3}, Fabrice Gravelat^{2,3}, Anne Moné^{2,3}, Corinne Biderre-Petit^{2,3}, Delphine Boucher^{1,2}, Julien Troquet⁴, Pierre Peyret^{1,2*}

Abstract

Background: Microorganisms display vast diversity, and each one has its own set of genes, cell components and metabolic reactions. To assess their huge unexploited metabolic potential in different ecosystems, we need high throughput tools, such as functional microarrays, that allow the simultaneous analysis of thousands of genes. However, most classical functional microarrays use specific probes that monitor only known sequences, and so fail to cover the full microbial gene diversity present in complex environments. We have thus developed an algorithm, implemented in the user-friendly program Metabolic Design, to design efficient explorative probes.

Results: First we have validated our approach by studying eight enzymes involved in the degradation of polycyclic aromatic hydrocarbons from the model strain *Sphingomonas paucimobilis* sp. EPA505 using a designed microarray of 8,048 probes. As expected, microarray assays identified the targeted set of genes induced during biodegradation kinetics experiments with various pollutants. We have then confirmed the identity of these new genes by sequencing, and corroborated the quantitative discrimination of our microarray by quantitative real-time PCR. Finally, we have assessed metabolic capacities of microbial communities in soil contaminated with aromatic hydrocarbons. Results show that our probe design (sensitivity and explorative quality) can be used to study a complex environment efficiently.

Conclusions: We successfully use our microarray to detect gene expression encoding enzymes involved in polycyclic aromatic hydrocarbon degradation for the model strain. In addition, DNA microarray experiments performed on soil polluted by organic pollutants without prior sequence assumptions demonstrate high specificity and sensitivity for gene detection. Metabolic Design is thus a powerful, efficient tool that can be used to design explorative probes and monitor metabolic pathways in complex environments, and it may also be used to study any group of genes. The Metabolic Design software is freely available from the authors and can be downloaded and modified under general public license.

Background

Assessing the metabolic potential of microorganisms in variable ecosystems is a novel and stimulating challenge in biology. Microorganisms are present in all environmental habitats, even the most extreme, yet despite their ubiquity, we know relatively little about these communities. Microorganisms display vast diversity, each one having its own set of genes, cell components and

metabolic reactions [1]. Thus 1 g of soil may contain up to 10^9 bacteria cells, which may represent between 1,000 and 10,000 different species [2,3]. Assuming 3,000 genes per single bacteria genome, there will thus be up to 3×10^{12} genes mediating huge and various biological processes [3,4]. To overcome the limits of cultivation, several high throughput approaches have been developed to explore genetic contents, such as metagenomics or DNA microarrays [1,5,6]. Numerous random shotgun metagenomic projects have caused the publicly available sequence data to increase exponentially, giving us a basis to study complex ecosystems [1,5]. In some cases,

* Correspondence: pipeyret@univ-bpclermont.fr

¹Clermont Université, Université d'Auvergne, Laboratoire: Microorganismes Génome et Environnement, BP 10448, F-63000 CLERMONT-FERRAND, France
Full list of author information is available at the end of the article

these sequence data were used to identify different species in environmental or clinical samples with DNA microarrays [5]. Moreover, these data should improve our knowledge not only of genome organization and genome evolution but also of biological processes and biological activities. However, although such sequencing approaches can rapidly generate large amounts of data, they give only a snapshot of genetic information and can be laborious and costly when complex ecosystems are to be studied. Also, DNA sequencing is not informative on gene expression and regulation. Metatranscriptomic studies are promising, but several obstacles have to be crossed before they can be widely used [1,7]. Indeed, sequencing approaches highlight the difficulties of accurate functional annotation of unknown proteins without experimental data; unsupervised annotation of proteins by software pipelines suffers from very high error rates. Spurious functional assignments are usually caused by species homology-based transfer of information from existing database entries to new target sequences [8,9]. Such functional annotation errors are due to local similarities between the query and functionally annotated sequences. Hence, two protein sequences may have two different biological functions, but a same protein domain. This approach, based on homologous gene prediction, presents another major drawback: it can fail to identify novel enzymes that have the same function, but a different primary structure from known enzymes [10]. Today, the main sources for such protein sequence data are Swiss-Prot, TrEMBL and GenPept. This last should be considered as an equivalent to the Swiss-Prot/TrEMBL databases with a high level of redundancy in terms of protein sequences [11]. Unlike TrEMBL, the Swiss-Prot database contains curated datasets of high quality [12].

Another high throughput tool, functional DNA microarrays, can also be used for monitoring metabolic diversity of microbial populations in environmental samples. In a single experiment, thousands of genes can be simultaneously detected. Several studies already demonstrate the usefulness of functional DNA microarrays for exploring various ecosystems [13-15]. Hybridization of microarrays with mRNA targets permits low-cost, easy quantitative estimates of gene expression levels [16]. Monitoring environmental metabolic processes can be made more powerful, and so more useful, by designing explorative probes to ensure the detection of genes not already discovered and deposited in databases. However, microarray probe design software determines specific probes to monitor only known sequences [17]. Thus only a small fraction of genes encoding microbial enzymes can be studied with these probes. To solve this problem, degenerate probes need to be defined, as for PCR-based applications [18].

Probe design also has to allow for the constraints of cross-hybridization. Specificity is a measure of the inability of a probe to bind strongly to non-target sequences that may be present in a biological sample. This can be accomplished by avoiding probes with excessive sequence similarity to a non-target sequence that may be present during the hybridization [19,20]. These problems of cross-hybridization emphasize the need to take into account the fact that the studies are conducted on complex environments. As thermodynamic constraints are not yet completely understood [21], sequence similarity is currently the prime parameter used to check probe specificity. A previously reported and extensively cited work by Kane and coworkers [22] on 50-mer probes, shows that a probe must meet two conditions to be specific: (i) the oligonucleotide sequence must have no more than 75% similarity (among all sequences) with a non-targeted sequence present in the hybridization pool, and (ii) the oligonucleotide sequence must not include a stretch of identical sequence longer than 15 contiguous bases.

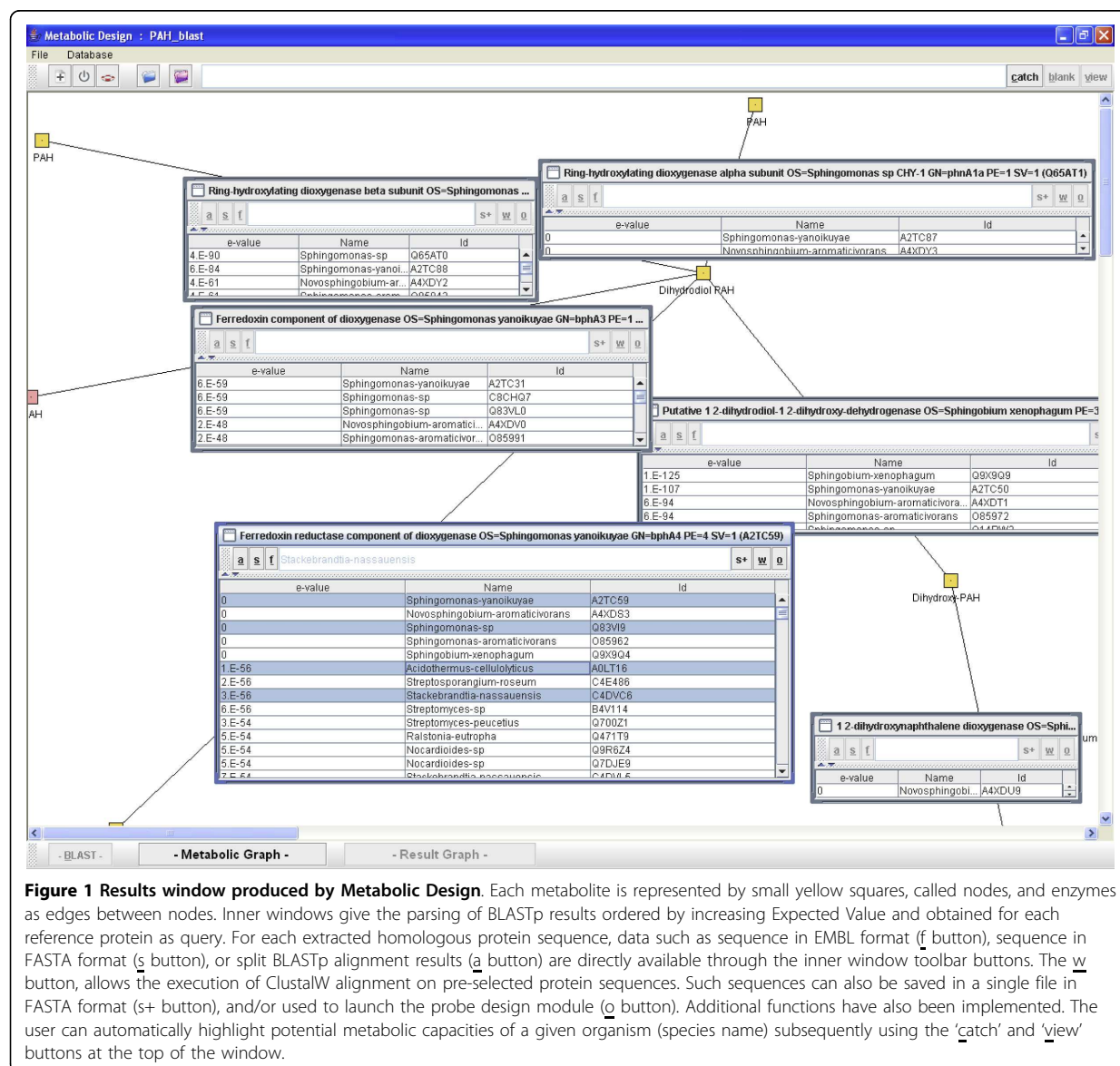
Here we describe a new algorithm, implemented in a user-friendly program, named Metabolic Design, which will generate efficient explorative probes using a simple convenient graphical interface. The practical utility of this approach was demonstrated by studying several genes encoding enzymes involved in the degradation of diverse polycyclic aromatic hydrocarbons from the model strain *Sphingomonas paucimobilis* sp. EPA505 (strain EPA505) and assessing metabolic capacities of microbial communities in a soil contaminated with aromatic hydrocarbons.

Results

The Metabolic Design software

Our aim is to build a graphical display of given biological processes and perform exhaustive sequence mining of all available protein sequences for each biological step studied. The graphical user interface (GUI) allows for example the graphical reconstruction of tailor-made metabolic pathways, with metabolites and enzymes represented respectively with nodes and edges (Figure 1). Using appropriate keywords, correctly annotated protein sequences are extracted from a curated database (by default Swiss-Prot potentially enriched with personal data) for each edge of the graph. The user can freely select the most suitable protein as a reference sequence query. This sequence is then used to carry out exhaustive mining of similar proteins from public and/or personal databases. The strategy of probe design using Metabolic Design software is described in Figure 2 and detailed in Methods under 'Software implementation'.

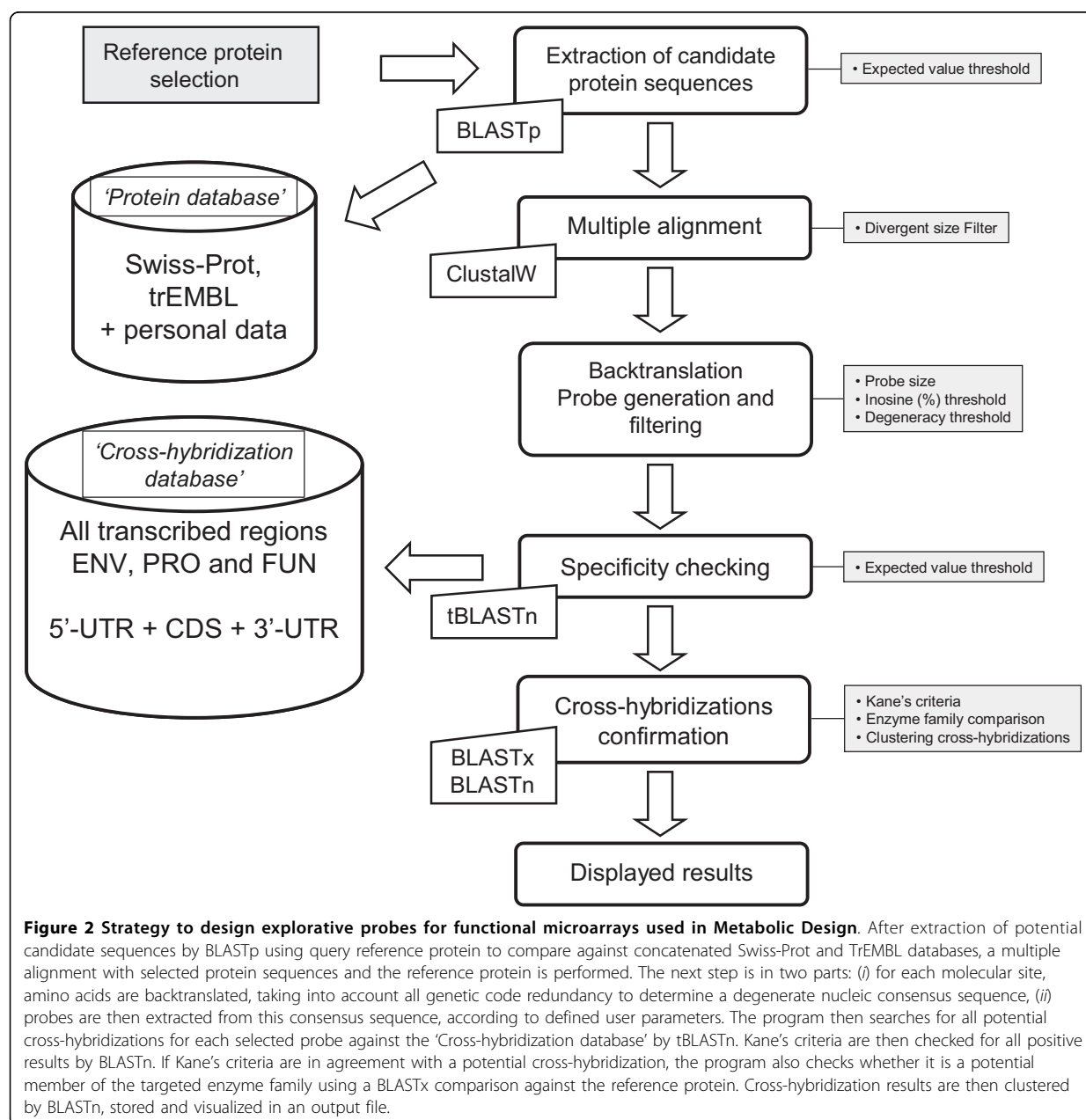
In our study, reference sequences are extracted from the highly curated database Swiss-Prot formatted for the



application to ensure efficient mining. In addition, the reference database is enriched with TrEMBL protein sequences biologically validated when a non-studied orthologous protein sequence was found in the Swiss-Prot database. For every listed protein, data are extracted using a homologous approach with a BLASTp program against concatenated Swiss-Prot and TrEMBL databases. Thus the selection of candidate protein sequences based on similarity criteria bypasses functional annotation errors. Extracted sequences are then automatically filtered and displayed in graphical edges for each studied enzyme. The results are finally organized according to increasing expected value, or organism origin, and miscellaneous functions are also

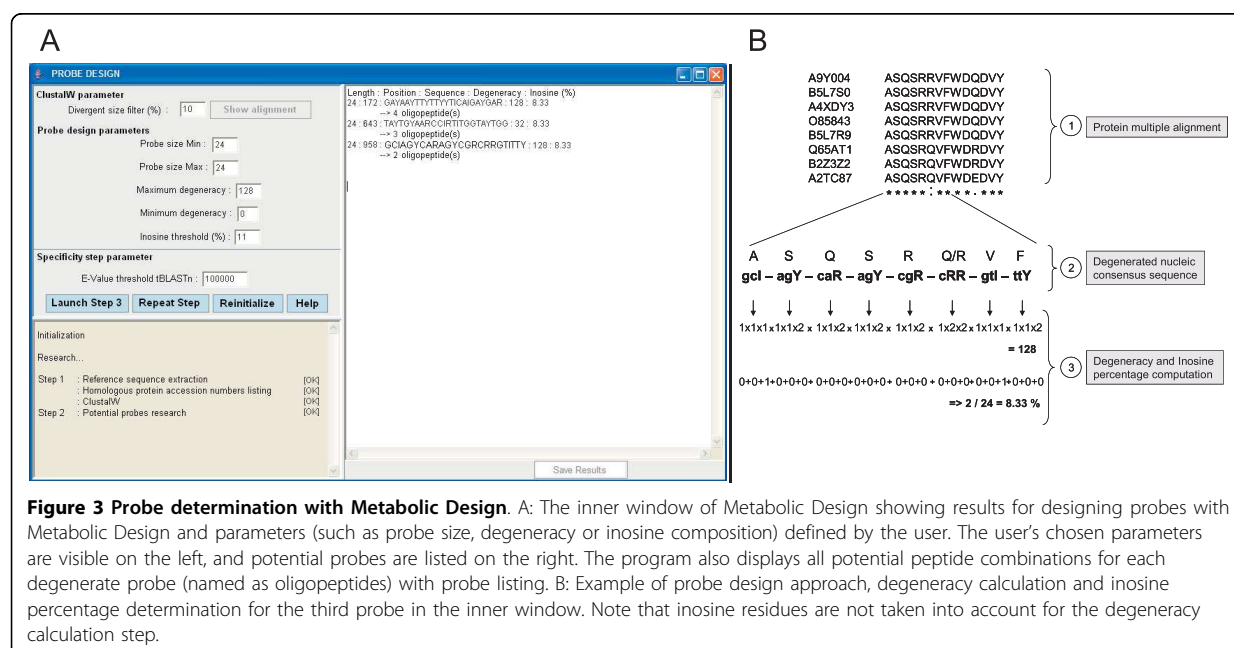
implemented in the toolbar to facilitate additional data extraction and visualization (Figure 1).

This multiple alignment is then used to design specific explorative oligonucleotide probes targeting studied proteins (Figure 3A), using the following procedure. To reduce insertion-deletion (indel) regions in multiple alignments, a first filtering step is carried out to exclude sequences with high size divergence compared with the reference query. A degenerate nucleic consensus sequence based on the IUPAC (International Union of Pure and Applied Chemistry) nomenclature is defined from the protein multiple alignment using the backtranslation approach [23]. For each molecular site, potential amino acids are backtranslated taking into account all



genetic code redundancy (Figure 3B). Probes are then extracted from this consensus sequence, according to three defined user parameters: probe size, degeneracy and inosine composition thresholds. Along the consensus sequence, the algorithm extracts all probes by incrementing the constant defined probe size in a window. All probes with degeneracy and inosine composition under the set thresholds are then listed in an inner window in the GUI. Thus the user can select all or some pre-selected potential probes for specificity testing. To reduce computing time, during this test the algorithm

generates all peptide combinations for each degenerate probe. Thus the degeneracy code redundancy is bypassed and the number of comparisons is greatly reduced. This test is carried out using tBLASTn against the 'Cross-hybridization database' using Kane's algorithm criteria [22]. Indeed, those parameters are used to check all positive results by comparison at the nucleotide level with BLASTn. If those criteria are in agreement with a potential cross-hybridization this may also reflect hybridization with a member of the targeted enzyme family. To avoid this bias, the algorithm extracts



the complete sequence of the gene harboring the potential cross-hybridization region and compares it with the reference protein using the BLASTx program. Finally, a file containing all potential cross-hybridizations for every candidate probe is automatically clustered, created and displayed.

Data mining and probe selection for microarray experiments using Metabolic Design

To validate our probe design strategy, we focus on metabolic pathways involved in the biodegradation of polycyclic aromatic hydrocarbons (PAHs). PAHs are a class of fused-ring aromatic compounds that are ubiquitous environmental pollutants known to be toxic, mutagenic and/or carcinogenic. Many researchers have therefore focused on the biodegradation of these pollutants by microorganisms, especially bacteria. Several enzymes involved in these critical biodegradation steps have been characterized and their sequences deposited in databases [24–28].

In this study, we choose to target eight genes (*phnA1a*, *phnA2a*, *bphC*, *bphA3*, *ahdA1c*, *ahdA2c*, *ahdA4* and *bphB*) (Table 1) known to be involved in the degradation of several PAHs (such as phenanthrene (PHE) and fluoranthene (FLA)). Using our defined data mining strategy, we first construct the metabolic pathway with respective substrates and products of each metabolic step. Secondly, for each of these metabolic steps, one reference enzyme is extracted from our curated database (Swiss-Prot and trEMBL validated data). Homologous proteins are then retrieved from

complete databases (Swiss-Prot and TrEMBL). Based on defined expected threshold values, different sequences are selected (Table 1) and multiple alignments are then performed to ensure probe design step.

To improve our probe design, we have applied two different strategies, using the same multiple alignments. In these strategies, we set the probe length at 24-mer, representing the best compromise between probe specificity and sensitivity criteria [29]. In the first strategy, (degeneracy threshold: 129, inosine threshold: 25%), we have determined a first set of probes for each targeted enzyme. However, owing to high percentages of inosine, these probes generally show a high degree of total degeneracy. Indeed, like the inosine residues are not taken into account for the degeneracy threshold determination, probes may present a maximum total degeneracy of 528,384. To reduce the number of potential specific probes derived from each degenerate defined probe, a second strategy with more stringent parameters is applied (degeneracy threshold: 258, inosine threshold: 9%, maximum total degeneracy of 4 128). Using these parameters, we have found another set of probes for each targeted enzyme. We then choose among the two probe sets obtained with these two strategies, the best probes based on several sequentially evaluated criteria: (1) the total number of potential cross-hybridizations to decrease possibilities of non-homologous hybridizations, (2) the probe total degeneracy (including inosine composition) to restrict the number of specific probes in the microarray, and (3) the position of each probe in the reference sequence to target different regions for each

Table 1 Reference enzyme information

Gene	Enzyme	REFERENCE PROTEIN			BLAST THRESHOLD AND SEQUENCES USED	
		Organism	Accession Number	Reference	BLASTp e-value	Chosen enzymes for the probe design
<i>phnA1a</i>	Putative alpha subunit of ring-hydroxylating dioxygenase	<i>Sphingomonas</i> sp. CHY-1	Q65AT1	[24]	1e-40	B2Z3Z2, A2TC87, A4XDY3, 085843, A9Y004, B5L7S0, B5L7R9, Q1HCP6, Q7WUA0
<i>phnA2a</i>	Putative beta subunit of ring-hydroxylating dioxygenase	<i>Sphingomonas</i> sp. CHY-1	Q65AT0	[24]	1e-30	A2TC88, A4XDY2, 085842, B5L7R8
<i>ahdA1c</i>	Putative large subunit of oxygenase	<i>Sphingomonas</i> sp. P2	Q83VL2	[27,32]	1e-40	A2TC29, A9XZZ2, Q65AS5
<i>ahdA2c</i>	Putative small subunit of oxygenase	<i>Sphingomonas</i> sp. P2	Q83VL1	[27,32]	1e-30	A9XZZ3, Q65AS6, A4XDV1, 085992, A2TC30, Q9Z4T6
<i>bphB</i>	Putative 1, 2-dihydrodiol-1, 2-dihydroxy-dehydrogenase	<i>Sphingobium xenophagum</i>	Q9X9Q9	[25]	1e-40	Q14RW3, 085972
<i>bphC</i>	Putative biphenyl-2,3-diol 1,2-dioxygenase	<i>Sphingobium xenophagum</i>	P74836	[25]	1e-40	PI 1122, Q6LCU9, Q7DG81, A4XDU9, 085990, A9XZZ5, Q65AS8, Q9KW12
<i>bphA3</i>	Putative ferredoxin component of dioxygenase	<i>Sphingomonas yanoikuyae</i>	A2TC31	[24,32]	1e-20	034128, Q65AS7, A9XZZ4, A4XDV0, 085991, Q83VL0
<i>ahdA4</i>	Putative ferredoxin reductase component of dioxygenase	<i>Sphingomonas yanoikuyae</i>	A2TC59	[24,32]	1e-40	Q83VI9, A4XDS3, 085962

Organism name and source, accession number and bibliographic reference for each reference protein. BLASTp expected threshold values used and selected sequences for multiple alignments for the probe design are given.

enzyme. Also, to reduce the number of specific probes synthesized on the microarray, the last nucleotide of each probe (generally a degenerate base or an inosine due to degeneracy of the genetic code) is also manually eliminated.

Thus by these strategies, two degenerate probes targeting two different regions are selected per targeted

gene (Table 2). Based on these sixteen 23-mer degenerate probes, we finally obtain 8,048 specific probes.

Explorative probe validation

Strain EPA505 is known to utilize PHE and FLA as sole sources of carbon and energy for growth [30]. However, for this strain, the enzymes involved in the catabolism

Table 2 Selected probe information

Targeted Gene	Probe name	Sequence	Number of unique DNA sequences used for the probe design	Number of specific probes	Positions on the reference gene sequence
<i>phnA1a</i>	phnA1a_MD_A	GTITGYAAAYTAYCAYGGITGGGT	5	256	294 - 316
	phnA1a_MD_B	CAYGARATHGARGTITGGACITA	4	384	957 - 979
<i>phnA2a</i>	phnA2a_MD_A	GARGAYATHCAYTAYTGGATGCC	2	48	123 - 145
	phnA2a_MD_B	GGICARGTITGGATGGARGAYCC	3	128	261 - 284
<i>ahdA1c</i>	ahdA1c_MD_A	GARTGYGTITAYCAYCARTGGGC	3	128	318 - 340
	ahdA1c_MD_B	GAYGCIGCIGAYAARCARGCITA	2	1024	771 - 793
<i>ahdA2c</i>	ahdA2c_MD_A	GAYGAYMGIYTIGARGARTGGCC	3	1024	081 - 103
	ahdA2c_MD_B	ATHGAYACIATGATGGTIMGICC	3	768	459 - 481
<i>bphB</i>	bphB_MD_A	AAYGTIGGIATHTGGGAYTWYAT	3	768	261 - 283
	bphB_MD_B	AAYBTIAARGGITAYTTYTYGG	3	384	348 - 370
<i>bphC</i>	bphC_MD_A	CCITAYTTYATGCAYTGAAAYGA	5	128	558 - 580
	bphC_MD_B	TGGYTITGGGARTTYGGITGGGG	4	128	777 - 799
<i>bphA3</i>	bphA3_MD_A	ATHATHGARTGYCCITTYCAYGG	2	576	180 - 202
	bphA3_MD_B	ATHGAIGAYGGITGGTITGYAT	3	768	279 - 302
<i>ahdA4</i>	ahdA4_MD_A	GCIAAYGTICCIAGAAAYTTYTT	2	1024	159 - 181
	ahdA4_MD_B	CARGARACITAYCARAAAYGCIGC	2	512	867 - 889

Total number of specific probes from the probe degenerate sequence and relative positions on the reference gene sequence for each targeted gene are described. Numbers of unique DNA sequences, coding for studied enzymes are also given to highlight that our probes target known genes but also unknown ones. **Nomenclature:** M: A and C; R: A and G; W: A and T; S: G and C; Y: C and T; H: A, C and T; D: A, G and T; B: G, T and C; I: A, C, G and T.

of PHE and FLA have not been fully characterized. Only gene fragments for the ferredoxin component of dioxygenase (*pbbB* equivalent to *bphA3*) and for the 1,2-dihydroxy-biphenyl-2,3-diol 1,2-dioxygenase (*pbbA* equivalent to *bphC*) are available in public databases [31] for the studied enzymes. This strain is thus an excellent model to validate our approach, as we could work with no prior assumptions using explorative probes to ensure the detection of unidentified genes.

With this aim, growth kinetics experiments with PHE, FLA and a mix of both pollutants as sole carbon and energy source are carried out to evaluate the targeted gene expression. As expected, for the eight genes studied, we have detected positive hybridizations ($\text{SNR}' > 3$) on the DNA microarray using mRNA as targets extracted after 3 h of culture (Table 3). Surprisingly, we do not observe positive signals with the probes targeting one region of the *phnA2a* gene. However, one probe targeting the second region of this gene allow the detection of strong hybridization signals ($\text{SNR}' = 22.64 \pm 3.21$) indicating a potentially high level of gene expression induced by PAHs. Additionally, control experiments with glucose as sole carbon and energy source do not give positive hybridizations for most of the targeted genes (Table 3). The SNR' value indicating positive hybridization is close to the threshold reflecting a low gene expression. These results suggest that all the studied genes can be induced in response to the mix of PAH exposure. The same results are obtained for growth kinetics with one PAH (PHE or FLA) as sole carbon and energy source. The same specific probes give the highest SNR' for the eight targeted genes, but with different levels of induction. For example, for the same specific probe (named *bphA3_MD_B_0333*) targeting the region B of the gene *bphA3* in all PAH-cultures we find: 9.79 ± 1.39 with a mixture of two pollutants, 20.00 ± 5.84 with PHE alone, 7.50 ± 2.03 with FLA

alone and no positive signal with glucose. We note that the number of probes giving a positive signal is low for targeted genes (between 1 for *phnA2a* and 5 for *bphB* after 3 h of culture with the mix of PAHs) reflecting variable levels of similarity between targets and probes deduced from variably degenerate regions.

Based on these results, we can also predict the most likely gene sequence of the targets interacting with probes. Among the positive probes, one shows a strong signal (e.g. one targeting *bphA3* with a median $\text{SNR}' = 36.87 \pm 7.83$) compared with the others targeting the same region. We hypothesize that the strongest SNR' probe perfectly matched, or is the closest sequence to targeted genes. Using sequences of *bphA3* and *bphC* genes available in databases [EMBL: AF259397 and AF259398], we demonstrate that only two probes among the four have identical sequences with *bphC* and *bphA3* genes. These data do not confirm the efficiency of our approach, and so to validate our first observations, we decide to isolate and characterize these genes and the others by a combination of amplification, cloning and sequencing strategies. Four gene clusters of 4.47 kb, 2.13 kb, 1.20 kb, and 0.32 kb, respectively [EMBL: FM882255, FM882254, FM882253 and FN552592] are thereby obtained. The complete nucleotide sequence of the 4.47 kb contig [EMBL: FM882255] shows six putative non-overlapping open reading frames (ORFs). Among these, four are targeted with our microarray probes. The first encodes a polypeptide 98% similar to a putative biphenyl-2,3-diol 1,2-dioxygenase known to degrade various dihydroxy-PAHs, and named BphC [EMBL: BAC65429]. The second encodes a polypeptide 90% similar to a putative ferredoxin component of dioxygenase, named BphA3 [EMBL: BAC65428], involved in various steps of the process of PAH degradation for the electron transfer from reductase to dioxygenase complex [26]. Interestingly, these two ORFs are highly similar to

Table 3 Results obtained with designed probes for a mixture of phenanthrene and fluoranthene.

Gene name	<i>phnA1a</i>		<i>phnA2a</i>		<i>ahdA1c</i>		<i>ahdA2c</i>		<i>bphB</i>		<i>bphC</i>		<i>bphA3</i>		<i>ahdA4</i>	
Targeted region	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Total number of specific probes	256	384	48	128	128	1024	1024	768	768	384	128	128	576	768	1024	512
Number of specific probes giving a positive signal ($\text{SNR}' > 3$)	1	2	0	1	3	1	2	1	4	1	1	1	3	1	0	0
Highest median SNR' obtained for each targeted region	18.32 \pm 3.64	6.62 \pm 0.31	X	22.64 \pm 3.21	8.61 \pm 1.59	9.93 \pm 1.32	8.92 \pm 1.52	16.26 \pm 2.45	5.79 \pm 1.73	4.09 \pm 0.66	4.47 \pm 0.30	4.54 \pm 0.81	36.87 \pm 7.83	9.79 \pm 1.39	X	X
Specific probe for EPA505 gene giving highest median SNR'	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No
For comparison, total number of specific probes giving a positive signal with glucose	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

For each degenerate probe defined targeting two different regions (A and B) of genes (*phnA1a*, *phnA2a*, *ahdA1c*, *ahdA2c*, *bphB*, *bphC*, *bphA3* and *ahdA4*), total number of specific probes stemming from the degenerate sequence, total number of specific probes giving a 'positive' signal (with a $\text{SNR}' > 3$), highest median SNR' visualized for each targeted region of each gene and whether the probe specific to the strain EPA505 gene gives this highest signal median SNR' .

available sequences for strain EPA505 [31], but a comparison with our sequences reveals some mismatches. The two last genes encode two polypeptides respectively 88% and 95% similar to AhdA2c [EMBL: BAC65427] and AhdA1c [EMBL: BAC65426], two components of a terminal oxygenase involved in the monooxygenation of salicylate, a metabolic intermediate of PHE, to catechol [32,33]. Two genes identified on the 2.13 kb contig (FM882254) encode polypeptides of 455 and 175 residues. These polypeptides resemble in length and sequence the alpha (99% sequence identity) and beta (100% sequence identity) subunits [EMBL: CAG17576 and CAG17577] of the ring-hydroxylating dioxygenase (*phnA1a* and *phn2a* respectively) of *Sphingomonas* sp. CHY-1, involved in the conversion of several PAHs into their corresponding dihydrodiols [28,34]. The third contig of 1.20 kb (FM882253) encompasses a single partial ORF encoding a polypeptide displaying 95% similarity with the ferredoxin reductase component of a dioxygenase, named AhdA4 [EMBL: BAC65450] of *Sphingobium* sp. P2 and involved in the electron transfer in association with BphA3 [35]. The last contig of 0.32 kb [EMBL: FN552592] encodes a partial 107 amino acid sequence 97% similar to a 1,2-dihydrodiol-1,2-dihydroxy-dehydrogenase named BphB [EMBL: ABM79802] of *Sphingobium yanoikuyae* B1.

Comparison of these gene sequences with the microarray probes shows that our design strategy is efficient to detect, with no prior sequence assumptions, targeted genes from complete metabolic pathways. As expected, for each gene, different probes give positive signals in agreement with the gene sequence composition. Furthermore, among the thirteen probes (targeting both regions of the eight genes) giving the highest signals, nine probes perfectly match strain EPA505 targeted gene regions (Table 3). Thus the two regions (A and B) selected for *bphA3* and *ahdA2c* genes probe designs allow the specific identification of these genes. For the genes *phnA1a*, *phnA2a*, *ahdA1c*, *bphB* and *bphC*, only one region can be considered specific for the identification of the genes. Finally, for *ahdA4* gene, as no probes give positive signals, we can then hypothesize that *ahdA4* is not expressed or is weakly expressed (under the detection threshold) in our culture conditions. We can also postulate that absence of signal might reflect a low sensitivity of these selected probes targeting *ahdA4*.

To conclude, these results confirm that our design strategy is useful and efficient for the targeted genes studied. These data also show that it is essential to select at least two specific regions for each studied gene that should be experimentally validated to ensure accurate identification. Nevertheless, a majority of selected regions is useful for the design of efficient probes that

perfectly hybridize with their targets and show the strongest signal on the microarray.

Gene expression analysis with microarray and quantitative real-time PCR experiments

As described previously, the applied design strategy lets us to detect targeted genes from the studied metabolic pathway without prior assumptions. It is thus of interest to test whether our DNA microarray is able to evaluate mRNA levels semi-quantitatively during biodegradation kinetics with PHE, FLA and a mixture of the two pollutants as sole carbon and energy source. A control experiment with glucose as sole carbon and energy source is also conducted. For these four conditions, total RNAs are extracted from pure cultures of strain EPA505 at different times of the kinetics (0, 3, 6, 10 and 21 h). According to the explorative probe validation conclusions (see previous section), only the most efficient probes targeting each of the eight genes in response to pollutant exposure are considered. In addition, to evaluate the gene expression level, a quantitative reverse transcription PCR approach is also developed for the selected genes during the same times of the kinetics.

Transcript hybridizations obtained with only glucose-amended cultures give no positive probe signals ($\text{SNR}' > 3$) for the different times of the kinetics studied as shown in Additional file 1. Under PHE-growth conditions, specific probes give positive signals ($\text{SNR}' > 3$) after 3 h of growth for all the studied genes (Additional file 1). Detected signals largely decrease at 6 h of culture to reach SNR' values under the set threshold. Same SNR' values, in agreement with absence or low abundance of targeted mRNA, are also obtained after 10 h and 21 h of culture (Additional file 1). With FLA as carbon source, except for *ahdA1c*, *bphC* and *bphB* probes, positive SNR' values are also obtained with specific probes after 3 h of growth. After 6 h of culture with FLA, no positive probe signal ($\text{SNR}' > 3$) is detected, as in glucose-growth conditions (Additional file 1). Surprisingly, a positive signal for the specific probe targeting *bphB* is detected after 6 h of culture ($\text{SNR}' = 3.43 \pm 0.70$) with FLA, but not after 3 h of culture. Finally, with a mixture of the two pollutants, high positive signals are detected, except for the *ahdA4* gene, under the SNR' threshold and for *bphC* and *bphB*, just above the SNR' threshold, after 3 h of culture (Additional file 1). After a large decrease in SNR' values after 6 h of culture, positive signals for most of the probes are visualized after 10 h of culture, indicating a new gene expression induction. Finally, at 21 h of culture, the detected signals have the same SNR' values as those obtained with glucose. Gene expression results obtained with microarray assays show an up-regulation of all the studied genes with different mRNA levels according to

PAH exposure (Additional file 1). For *ahdA4*, no positive signals are detected except with PHE after 3 h of culture with a SNR' close to the threshold ($\text{SNR}' = 3.19 \pm 0.40$).

At the same time, a quantitative reverse transcription PCR based approach is used to precisely describe the gene expression during the growth kinetics. Results show the same expression profiles as those observed with DNA microarray experiments (Additional file 1). Low mRNA levels are detected during growth on glucose, indicating a very low basal gene expression in the absence of PAH substrates. With PHE or FLA as sole carbon and energy source, a high level of targeted mRNA is detected after 3 h of growth. However, a higher mRNA level is detected with PHE exposure. For these two cultures, after 10 h of culture, gene transcript number decreases to reach mRNA levels close to or below the control copy number detected in glucose-grown cells, as with results visualized with the DNA microarrays. With a mixture of the two pollutants, the same expression profile is detected with the quantitative reverse transcription PCR approach and with the DNA microarrays. High mRNA levels are measured after 3 h of culture, and besides a large decrease after 6 h of culture, another mRNA up-regulation is detected at 10 h of culture for the studied genes. Finally, mRNA levels decrease to reach transcript levels close to growth experiments performed with glucose. In conclusion, similar expression profiles are obtained for *phnA1a*, *phnA2a*, *ahdA1c*, *ahdA2c*, *bphB*, *bphC* and *bphA3* with DNA microarray and quantitative reverse transcription PCR approaches, demonstrating the efficiency of probes designed using Metabolic Design software. Thus DNA microarrays using Metabolic Design can be used to perform semi-quantitative monitoring of gene expression.

Characterization of potential metabolic capacities in a PAH contaminated soil

As we developed explorative probes to detect key genes coding for enzymes involved in PAH degradation, we assess the metabolic capacities of endogenous microbial communities in a polluted ecosystem. Owing to the difficulty in extracting microbial RNA in such environments, we hybridize total extracted microbial DNA from a highly contaminated soil (contamination details in Additional file 2). This ecosystem is selected because it harbors high concentrations of PAHs (2,300 mg/kg of dry soil). Also, PHE and FLA are detected as major contaminants (respectively 430 and 270 mg/kg of dry soil).

Among the 8,048 designed probes targeting the eight genes, 358 give positive signals ($\text{SNR}' > 3$) after hybridization with total DNA (Table 4). For each gene, probe sets show strong signals, but with variable intensities, identifying the most probable target sequence. To

evaluate the explorative capacities of our probes, we first focus on the *phnA2a* gene. We compare the signal intensities between mRNA hybridization of strain EPA505 and the DNA extract from the polluted soil (Figure 4). We clearly identify the probe signature for strain EPA505 and a specific probe signature for the polluted soil. Using a BLASTn approach with complete databases (EMBL), 21 positive probe sequences have high similarities (0, 1 or 2 mismatches) with *phnA2a* genes from known PAH degraders (such as *Novosphingobium* sp. H25, *Cycloclasticus* sp. NY93E or *Sphingomonas* sp. CHY-1) (data not shown). We can then hypothesize that other positive probe sequences presenting a slight homology with available *phnA2a* sequences might have targeted *phnA2a* unknown genes, consistent with the explorative purpose of these probes.

The highest SNR' signal is given for a probe targeting *ahdA1c* (42.85 ± 5.83) among 204 other positive probes for this gene. As for *phnA2a* positive probes, several are potentially explorative. Interestingly, specific probe targeting *ahdA1c* gene from strain EPA505 also gives a positive signal (median $\text{SNR}' = 7.45 \pm 0.34$). The same positive results are obtained with probes specific to strain EPA505 genes: 3.12 ± 1.00 for *phnA2a*, 4.07 ± 0.27 for *ahdA2c*, 4.33 ± 1.14 for *bphC* and 7.06 ± 1.22 for *bphA3*, suggesting the presence of bacteria closely related to strain EPA505.

Surprisingly, no probe can detect *phnA1a* gene in the polluted soil. We choose to amplify, with a PCR approach, *phnA1a* genes using degenerate primers (data not shown). The PCR products are then cloned, and eight clones are sequenced. Among these eight sequences, seven showing high similarities with *phnA1a* genes are then compared with our probe sequences. This comparison reveals multiple mismatches (data not shown), impeding hybridizations with our probes. This result indicates a marked divergence of this gene family. Our first design focused on *phnA1a* genes related to *Sphingomonas*. For a broader discovery of gene diversity, we will need to design probes that take into account more exhaustively the most complete sequence diversity in databases (international and/or personal).

Discussion

We have developed and validated a new algorithm named Metabolic Design. This software can be used to design efficient explorative probes for functional DNA microarrays. Previously to probe design, users have to extract from public (Swiss-Prot and TrEMBL) or personal databases, protein sequences of interest. Results are then integrated in a user-friendly, intuitive interface. All databases used for the application can be selected by the users and they can also integrate personal data. Such flexibility is generally not available, for example with

Table 4 Results obtained with designed probes with total DNA extracted from the contaminated soil S3

Gene name	<i>phnA1a</i>	<i>phnA2a</i>	<i>ahdA1c</i>	<i>ahdA2c</i>	<i>bphB</i>	<i>bphC</i>	<i>bphA3</i>	<i>ahdA4</i>
Targeted region	A	B	B	A	B	A	B	A
Total number of specific probes	256	128	1024	1024	768	768	128	768
Number of specific probes giving a positive signal (SNR' > 3)	0	37	204	18	1	36	16	44
Percentage of probes giving a positive signal (SNR' > 3)	0	28.90	19.92	1.75	0.13	4.68	12.50	5.72
Highest median SNR' obtained for each targeted region	0 ± 0.00	9.47 ± 0.70	42.85 ± 5.83	7.05 ± 1.37	4.29 ± 1.71	6.33 ± 2.05	4.43 ± 1.31	8.84 ± 2.15
								0.98

For each degenerate probe defined targeting one particular region (A or B) of genes (*phnA1a*, *phnA2a*, *ahdA1c*, *ahdA2c*, *bphB*, *bphC*, *bphA3* and *ahdA4*), total number of specific probes stemming from the degenerate sequence, total number of specific probes giving a 'positive' signal (with a SNR' > 3), probe percentage giving a 'positive' signal and highest signal median SNR' visualized for each targeted region of each gene.

current metabolic reconstruction tools, such as the 'Pathway Tools Software', initially developed for the EcoCyc project [36], or KEGGanim [37]. These are generally based on static databases and predefined metabolic pathways (such as KEGG [38], MetaCyc [39] or BRENDA [40]).

In order to bypass the faulty annotations found in automatically filled databases, and to allow the exhaustive exploitation of all the currently available protein sequences, the mining step is performed using similarity search. However, such approach presents another major drawback. Indeed, in some cases, not all proteins with a similar function have similar primary structures. Thus a future development of Metabolic Design will be the replacement of the BLASTp step by a Pattern Hit Initiated BLAST (or PHI-BLAST) step coupled with PRODOM data (defined as a comprehensive set of protein domain families automatically generated from the UniProt Knowledge Database) [41]. PHI-BLAST analysis is useful for identifying the distant members of a protein family, whose relationship is not recognizable by straight

sequence comparison, but only by patterns contained in sequences (such as catalytic sites or substrate recognition sites). We also intend to integrate a new module for high-throughput ortholog prediction (using for example Ortho-MCL or Orthologue) to improve homologous protein selection for complex and divergent protein families [42,43].

The ultimate aim of Metabolic Design is to define explorative probes and estimate their specificity *in silico*. Specific probes deduced from defined degenerate probes thus allow the targeting not only of known gene sequences but also of new ones that encode the same protein sequences. These explorative features are not offered by other tools such as OligoArray 2.0, YODA or HPD [17]. In addition, Metabolic Design takes into consideration both *ex situ* and *in situ* DNA microarray synthesis. The inosine composition is taken into account in the total degeneracy, as an *ex situ* microarray can hold inosine nucleotide probes, and/or degenerate probes in one spot, reducing probe degeneracy.

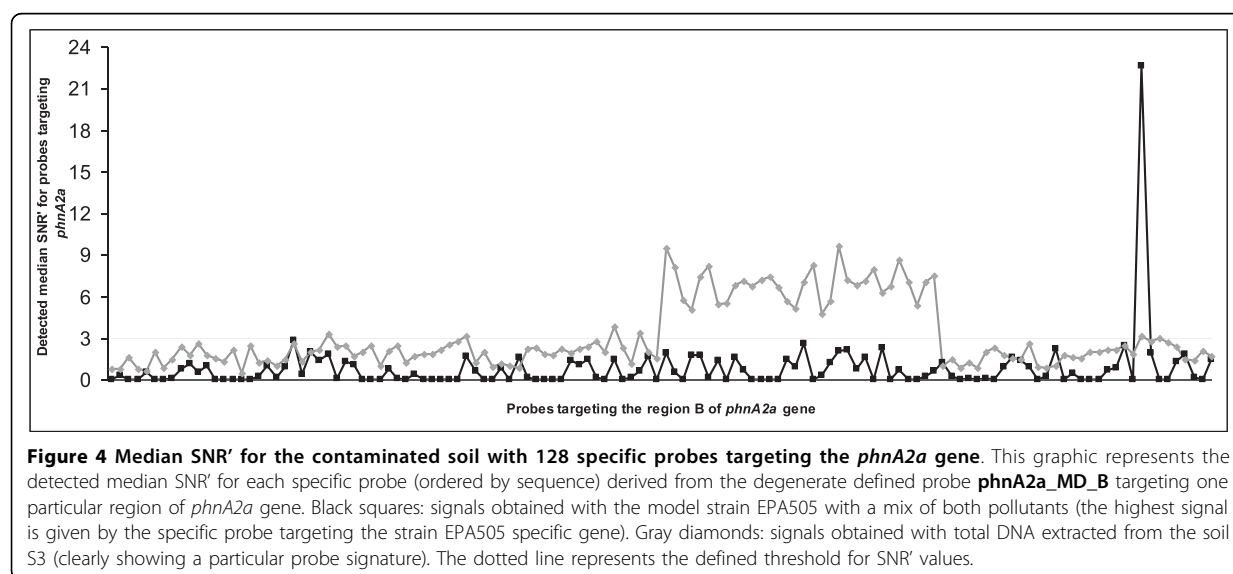


Figure 4 Median SNR' for the contaminated soil with 128 specific probes targeting the *phnA2a* gene. This graphic represents the detected median SNR' for each specific probe (ordered by sequence) derived from the degenerate defined probe *phnA2a_MD_B* targeting one particular region of *phnA2a* gene. Black squares: signals obtained with the model strain EPA505 with a mix of both pollutants (the highest signal is given by the specific probe targeting the strain EPA505 specific gene). Gray diamonds: signals obtained with total DNA extracted from the soil S3 (clearly showing a particular probe signature). The dotted line represents the defined threshold for SNR' values.

Probe specificity is then evaluated *in silico* using a proprietary database, giving us a close glimpse of potential cross-hybridizations found in complex environments. In addition, in Metabolic Design, this database can be modified to consider complete DNA data, or only fragmented data (for example, only one genome). Estimation and validation of potential cross-hybridizations are performed by a BLASTn analysis. However, one possible improvement would be to take into account optimized BLASTn parameters recently described as allowing a more efficient detection of potential cross-hybridizations [44].

Another update of Metabolic Design will add thermodynamic calculations to improve probe selection, although these parameters are not fully described at present [21,45]. Also, it will be essential to take into account probe sensitivity due to sequence nature [46]. In view of these difficulties in precisely predicting probe behavior during DNA microarray hybridizations, we suggest that users first validate the quality of the DNA microarrays (probe specificity and sensitivity), with a simple biological model as we did in this study.

Based on Metabolic Design defined probes, targeting eight genes coding for enzymes involved in the degradation of various PAHs by strain EPA505, we demonstrate that our design strategy is useful for most of the determined probes. Furthermore, these results highlight the capacity of our probes for semi-quantitative monitoring of gene expression or gene detection, confirming the quantitative capability of our microarrays for environmental applications [14]. Finally, we demonstrate the explorative ability of our probes, studying a complex environment. Indeed, most classical functional microarrays (such as GeoChip) using specific probes will monitor only known sequences and cannot appraise the complete microbial gene diversity of complex environments [13,14,47,48]. Additionally, considering the high complexity of environmental samples, it will be interesting to improve again probes specificity and sensitivity, using for example the 'GoArrays' strategy [29].

To allow the identification of complete sequences of targeted genes, a further application of these explorative DNA microarrays will be the capture of 'unknown' sequences for further next-generation sequencing [49,50]. Some new techniques have been reported for performing selective capture of sequence fragments from complex mixtures based on hybridization to DNA microarrays. Combining our explorative DNA microarrays and next-generation sequencing will, for example, bypass a critical bottleneck in microbial ecology, namely the difficulty of specifically exploring some biochemical pathways or specific biomarkers without the need to sequence the complete metagenome or PCR products (not reflecting reality due to PCR artifacts). Most often

in complex environments even with high throughput sequencing, we obtain only a partial view of the extremely broad microbial diversity. In addition, using mRNA or large DNA fragments as targets can allow all the genes included in a transcriptional unit to be captured. So, in prokaryotes, like genes involved in the same biological process are generally associated in the same transcriptional unit, this capture would allow to assign of new gene functions.

Conclusions

This study evaluates the efficiency of a new probe design software tool, Metabolic Design, dedicated to DNA functional microarrays. This software, which can be used to study any group of genes, was successfully applied to define probes able to detect with high specificity and sensitivity genes encoding enzymes involved in PAH degradation. In addition, DNA microarray experiments performed on soil polluted by organic pollutants, without prior sequence assumptions, demonstrate explorative abilities of our probes. So, probe design performed with Metabolic Design ensures to precisely monitor metabolic regulations during various processes in complex environments.

Methods

Software implementation

The Metabolic Design application can be obtained on request *via* FTP and runs only on MS-WINDOWS (32-bit) platforms. The Java runtime environment (JRE) Version 1.4 or higher, Perl Version 1.5 or higher and an SQL database such as Oracle 9i must be installed. Latest Swiss-Prot and TrEMBL database versions have also to be downloaded for local installation of data from ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete. Metabolic Design is a standalone multilayered tool comprising a relational database, a data object layer and a graphical user interface (GUI). For the data mining step, curated information associated with enzymes (systematic name and source organism) come from the international protein database Swiss-Prot/TrEMBL and/or personal data. Swiss-Prot/TrEMBL data are parsed using PERL scripts allowing extraction of files including sequences from prokaryotes and fungi. To evaluate potential cross-hybridization of candidate probes, microbial related sequences from the EMBL database, which include environmental samples (ENV), fungi (FUN) and prokaryote (PRO) taxonomic divisions, are selected. DNA sequences corresponding to CDSs with their respective putative 5' and 3' UTR flanking regions (arbitrarily set at 100 nt each) are then extracted and formatted to perform the cross-hybridization checking step by a tBLASTn approach ('cross-hybridization database'). For particular BLAST steps, some parameters

are defined for the Metabolic Design program: tBLASTn for the specificity step (*-e 10000000 -w 2 -b 5000 -v 5000*), BLASTx for the comparison with the reference protein sequence (*-e 1e⁻¹⁰ -F F*), BLASTn for Kane's criteria evaluation (*-e 10 -w 7 -F F -q -I*), and BLASTn for cross-hybridization clustering (*-e 10e⁻¹⁰ -w 7*).

A JAVA classes package is developed to implement the data object layer and GUI. Object persistence is guaranteed at both text file and SQL levels.

Chemicals

PHE, FLA, Tween 80, HPLC grade solvents and acetone are purchased from Sigma-Aldrich (Saint-Quentin-Fallavier, France). For degradation experiments, stock solutions of each PAH (PHE and FLA) are prepared in acetone at a final concentration of 2 g/L and sterilized as described above. A mixture of PHE and FLA (1 g/L each) is prepared in the same way.

DNA extraction from soil

Total DNA is extracted from 5 g of contaminated soil (S3) following the protocol described by Zhou [51]. Three extracts are made and pooled to minimize potential biases. DNA quality is checked on a 0.8% agarose gel.

Bacteria, growth conditions and kinetic experiments

Strain EPA505 (DSM7526) is purchased from DSMZ (Braunschweig, Germany). Cells are first grown overnight at 37°C on a shaker table (150 rpm) in 70 mL of Luria-Bertani medium (LB) containing streptomycin (100 mg/L) to produce biomass. They are then centrifuged at room temperature for 2 min at 5,000 g and transferred to a sterile minimum mineral medium 457 containing, for 1 L of distilled water, 2,440 mg of Na₂HPO₄, 1,520 mg of KH₂PO₄, 500 mg of (NH₄)₂SO₄, 200 mg of MgSO₄ · 7H₂O, 50 mg of CaCl₂ · 2H₂O, 200 mg of Tween 80 and 10 mL of SL-4 solution as described by DSMZ [30,33]. Cultures are prepared as follows: 2 mL of the PAH stock solutions is evaporated in sterile 250 mL conical flasks, 100 mL of sterile medium 457 is added, and the flasks are inoculated with 7.5 × 10⁷ cells prepared as above. These cultures are incubated at 28°C on a shaker table (150 rpm) for 27 h and bacterial growth is monitored spectrophotometrically at 620 nm using an Ultraspec 2000 spectrophotometer (Pharmacia Biotech AB, Uppsala, Sweden). A culture is also grown with glucose (15 g/L) as sole carbon source and energy to define the basal expression of genes implicated in PAH degradation.

RNA extraction from strain EPA505

Total RNA from a pure culture of strain EPA505 is extracted at different times of the PAH degradation

kinetics (0, 3, 6, 10, and 21 h) with the RNeasy Mini kit (Qiagen GmbH), and treated with 1.5 U of DNase I (Invitrogen) to eliminate DNA contamination. RNA sample concentration and purity are then estimated using a Nanodrop spectrophotometer (Nanodrop).

Microarray experiments

Samples of 15 µL of strain EPA505 total RNA of four PAH degradation kinetics data points (0, 3, 6, 10 and 21 h) are enriched using the MICROBExpress™ Bacterial mRNA Enrichment Kit (Ambion) as recommended by the suppliers. Each enriched mRNA is then amplified using the MessageAmp™ II-Bacteria RNA Amplification Kit (Ambion) with a modified protocol for the *in vitro* transcription step. Briefly, the purified double-stranded template (~14 µL) is transcribed *in vitro* with 12 µL of ATP, CTP and GTP mix (25 mM each) (Ambion), 3 µL of UTP (75 mM) (Ambion), 3 µL of amino-allyl-UTP (50 mM) (Ambion), 4 µL of 10 × reaction buffer (Ambion) and 4 µL of T7 enzyme mix (Ambion) at 37°C for a 14 h incubation period. Finally, the aRNA is purified using the MessageAmp™ II-Bacteria RNA Amplification Kit (Ambion) following the manufacturer's instructions.

In the next step, 10 µg of purified aRNA for each sample are vacuum-dried and labeled using the Amersham CyDye™ Post-Labeling Reactive Dye Packs (GE Healthcare, Little Chalfont, United Kingdom) with Cyanine3 or Cyanine5 dyes as recommended by the supplier. Briefly, the aRNA pellet is resuspended in 20 µL of 0.1 M bicarbonate buffer (pH 8.7) and incubated for 90 min with 40 nM of dye compound (coupling the dye to amino-allyl-UTP) dissolved in 20 µL of DMSO (dimethyl sulfoxide) in the dark at room temperature. Excess dye is quenched by adding 15 µL of 4 M hydroxylamine solution incubated for 15 min in the dark at room temperature. The labeled aRNA is then purified with NucleoSpin RNA Clean-Up kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. After each step (total RNA enrichment, RNA amplification and aRNA labeling), the quantity and integrity of RNA are estimated using the RNA 6000 Nano kit (Agilent Technologies), the Agilent 2100 Bioanalyzer (Agilent Technologies) and the Nanodrop spectrophotometer (Nanodrop) as recommended by protocols.

Total DNA is amplified and labeled using the Bio-Prime® Total Genomic Labeling System (Invitrogen) following the manufacturer's instructions. The quantity and quality of labeling are estimated using a Nanodrop spectrophotometer (Nanodrop) as recommended by protocols.

NimbleGen custom arrays of 8,048 probes are used (Roche NimbleGen, Madison, USA). All the probes

are randomly distributed across the array to minimize spatial effects as far as possible during the hybridization step. The microarray also contains thousands of random probes (randomly defined length and sequence) which can serve to measure technical background noise. For each hybridization experiment, 3.33 µg of labeled RNA for kinetic experiments (one sample in Cy3 and another in Cy5) or 12 µg of labeled DNA (in Alexa Fluor® 5) for soil are mixed, vacuum-dried and resuspended in 5.6 µL of water. The hybridization mix (Roche NimbleGen) is then made according to the manufacturer's protocols. The arrays are hybridized on a 4-bay NimbleGen Hybridization System (Roche NimbleGen) at 42°C for 72 h. The arrays are washed with NimbleGen wash buffers I, II and III according to vendors' protocols and scanned using a Scanner Innoscan 900AL (Innopsys, Carbonne, France) at 2 µm resolution. Individual array images are acquired independently, adjusting the PMT gain for each image as recommended using Mapix® software (Innopsys).

For each array image, raw expression data are extracted using the NimbleScan software v2.1. (Roche NimbleGen) and feature intensities are exported as .pair files. The background noise is then determined using random probes present on the microarrays (8,863 probes in our experiment) with the method described in the Additional file 3. This background noise is defined by two components: the background median intensity (Bposition) and its dispersion (Bdispersion). Finally, a modified signal-to-noise ratio termed SNR' and based on the formula of Verdik [52] is calculated as follows in order to reduce-centralize our data: $SNR' = (\text{probe signal intensity} - \text{Bposition}) / \text{Bdispersion}$ (see Supplementary Data S1).

However, spatial effect across the array surface is a predominant within-slide experimental artifact that needs to be eliminated before any other normalization procedure [53]. Accordingly, for all array images obtained in this work, the surface is segmented into 16 sub-squares according to probe position (X, Y) indicated in the pair report. A Perl script is developed to calculate local background noise in all sub-squares and the median SNR' retrieved from the three replicates of each probe. Finally, another Perl script is implemented to summarize each replicate probe treated and determine the median value of the three replicates. 'Positive' hybridization is considered significant for probes with $SNR' > 3$ (value to avoid all false positives) [54]. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE21402: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=rjmhhgoaqyqqinw&acc=GSE21402>

DNA extraction, PCR amplification and cloning

Total DNA from a pure culture of strain EPA505 is extracted by heat shocking cells [55]. All PCR reactions are carried out in 50 µL of mixtures containing 20 ng of the previous strain EPA505 DNA extract, 0.5 U of GoTaq DNA polymerase (Promega Corp., Madison, USA), Promega buffer, 1.25 mM MgCl₂, 1 µM of each primer depending on the targeted gene (Additional files 4 and 5) and 0.5 mM of each deoxynucleotide. The reactions are performed in a iCycler thermal cycler (Biorad Laboratories, Marnes-la-Coquette, France) using an initial denaturation step consisting of 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, an annealing step with temperature and time depending on primers (Additional Files 4 and 5) and an elongation step of 72°C for 1 min. A final treatment of 72°C for 7 min is then applied. The size and purity of PCR products are checked on 1.2% gel agarose. The PCR products are purified with a Qiaquick Gel Elution kit (Qiagen GmbH, Hilden, Germany), and then ligated into the pCRII-TOPO® vector supplied with the TA cloning kit (Invitrogen Corp., Merelbeke, Belgium) and transformed into *E. coli* One Shot® TOP10 cells (Invitrogen Corp.) following the manufacturer's instructions. White colonies are picked and grown in LB medium supplemented with kanamycin at 50 µg/ml final concentration. Plasmid template DNA is prepared by the alkaline lysis method [55]. The clone inserts are sequenced by the MWG Biotech Company (Ebersberg, Germany) using both SP6 and T7 sequencing primers. Sequence treatment and joining are performed using the pregap4 and the gap4 tools of the Staden Package Program [56]. The gene sequences are then compared with Swiss-Prot and TrEMBL databases using the BlastX program [57].

Real-time PCR experiments

The reverse transcription reactions are carried out at 42°C for 2 h with 50 ng of total RNA using *bphC*, *bphA3*, *ahdA2c* and *ahdA1c* (0.625 µM of each primer) mix primers, and *ahdA4*, *phnA1a*, *phnA2a* and *bphB* mix primers respectively (see Additional file 6) in order to minimize manipulation biases. These reactions are carried out in a final volume of 20 µL with 100 U of SuperScriptIII reverse transcriptase (Invitrogen Corp.), 1 U of RNasin+ Inhibitor (Promega Corp.), 0.25 mM of dNTPs mix (Invitrogen Corp.), 0.1 M DTT (Invitrogen Corp.) and Invitrogen buffer, according to the manufacturer's instructions. Reverse transcription reactions are performed in triplicate. cDNA is then diluted ten-fold for quantitative real-time PCR assays. Reactions are carried out with the MESA Green qPCR for SYBR assays kit (Eurogentec) according to the manufacturer's instructions. All amplifications are carried out in a final volume of 20 µL containing 5 µL of sample described above or 5

μL of standard cDNA (from 4.37×10^7 copies/μL to 4.37 copies/μL, covering 8 log of dynamic range for each gene), 10 μL of 2× MESA Green qPCR for SYBR assays mixture and the corresponding primers sets described in the Additional file 6 at 0.2 μM final concentration each. The reverse transcription product in kinetic experiment samples is quantified twice. As every reverse transcription experiment is done in triplicate, six measurements are obtained for each sample. Each point on the standard curve (corresponding to serially diluted cDNA) is quantified in triplicate. PCR is carried out in the Mastercycler Realplex (Eppendorf, Le Pecq, France) for 1 cycle at 95°C for 5 min followed by 40 cycles consisting of 95°C for 15 s (denaturation step) and 68°C for 45 s (annealing and elongation steps). At the end of the real-time PCR, a melting curve is defined by measurement of SYBR Green signal intensities for 20 min from 68°C to 95°C. Size of the amplified products is checked on 2.5% agarose gel. Data analysis is carried out with *realplex* software (version 1.5; Eppendorf).

Nucleotide sequence accession numbers

The nucleotide sequences reported in this study have been deposited in the database under accession numbers: [EMBL: FM882255] (encompassing *bphC*, *bphA3*, *ahdA2c* and *ahdA1c* gene sequences), [EMBL: FM882254] (encompassing *phnA1a* and *phnA2a* gene sequences), [EMBL: FM882253] (encompassing *ahdA4* gene sequences) and [EMBL: FN552592] (encompassing *bphB* gene sequences).

List of abbreviations

GUI: graphical user interface; IUPAC: international union of pure and applied chemistry; PAH: polycyclic aromatic hydrocarbon; PHE: phenanthrene; FLA: fluoranthene; SNR': signal to noise ratio; ORF: open reading frame; CDS: coding DNA sequence; RT-PCR: reverse transcription- polymerase chain reaction; FTP: file transfer protocol; JRE: java runtime environment; SQL: structured query language; UTR: untranslated region, aRNA: antisense RNA; DMSO: dimethyl sulfoxide; cDNA: complementary DNA.

Availability and Requirements

Project name: Metabolic Design

Project homepage: ftp://195.221.123.90/

Operating system: Windows (32-bit) only

Programming language: Java and Perl

Others: The Java runtime environment (JRE) Version 1.4 or higher, Perl Version 1.5 or higher and an SQL database such as Oracle 9i must be installed.

License: Free for non-commercial use. Source code available upon request.

Additional material

Additional file 1: SNR' profiles detected with microarray experiments and transcript numbers profiles detected with quantitative RT-PCR assays. SNR' profiles detected with microarray experiments (LEFT), and transcript copy number detected per ng of total RNA with quantitative RT-PCR assays (RIGHT) for eight genes: (A) *phnA1a*; (B) *phnA2a*; (C) *ahdA1c*; (D) *ahdA2c*; (E) *bphB*; (F) *bphC*; (G) *bphA3*; and *ahdA4* (H) during PAH biodegradation at different times with strain EPA505. PHE: grey squares, FLA: triangles, PHE + FLA: circles, glucose: open diamond. Error bars indicate the standard deviation of measures.

Additional file 2: PAH composition detected in the contaminated soil S3. These data are proprietary data given by BioBasic Environnement and give the quantity of detected PAHs in mg/kg of dry soil in the contaminated soil studied.

Additional file 3: Background noise calculation description. Background noise is determined according to 'RANDOM probes response' of Nimblegen microarrays. Our method takes into account the background noise which is characterized by two components: its position and its dispersion.

Additional file 4: Identification of four catabolic genes clusters from the model strain EPA505. Physical maps of four clusters (A, B, C and D) of catabolic genes involved in PAHs biodegradation from strain EPA505. Size of genes and intergenic spaces is indicated as well as position of primers used for PCR amplifications.

Additional file 5: Primer sets used for detecting catabolic genes involved in PAHs degradation and to generate the gene DNA matrix. The DNA matrix is used to build the standard curve for quantitative real-time PCR assays in strain EPA505. *: *xytX* and *nahD* are used to characterize complete sequences of *bphC* and *ahdA1c*.
Nomenclature: **M:** A and C; **R:** A and G; **W:** A and T; **S:** G and C; **Y:** C and T; **K:** G and T; **V:** A, G and C; **H:** A, C and T; **D:** A, G and T; **B:** G, T and C; **I:** A, C, G and T.

Additional file 6: Primers used for reverse transcription and quantitative real-time PCR assays. List of primers used for reverse transcription and subsequent quantitative real-time PCR assays. Amplification sizes are also given for each targeted gene.

Acknowledgements

We thank bioinformatics undergraduate students (IUT of Aurillac, Université d'Auvergne) Yann Keriou and Xavier Brotel for PERL script developments and Mathieu Roudel for FTP development, Mélanie Mitchell, an undergraduate student in biology (IUT of Clermont, Université d'Auvergne) for her help in quantitative reverse transcription PCR experimentation, and Brigitte Chebanne for technical assistance. ST was supported by a doctoral grant from 'Ministère de l'Enseignement Supérieur et de la Recherche Scientifique'.

Author details

¹Clermont Université, Université d'Auvergne, Laboratoire: Microorganismes Génome et Environnement, BP 10448, F-63000 CLERMONT-FERRAND, France. ²CNRS, UMR 6023, Laboratoire: Microorganismes Génome et Environnement, F-63173 AUBIERE, France. ³Clermont Université, Université Blaise Pascal, Laboratoire: Microorganismes Génome et Environnement, BP 10448, F-63000 CLERMONT-FERRAND, France. ⁴Biobasic Environnement, Biopôle Clermont-Limagne, 63360 Saint-Beauzire, France.

Authors' contributions

ST, EP, AM and OG carried out the experimentations and have participated to analysis and interpretation of data. FG, EP, ST and OG have made script developments, updates and optimizations of Metabolic Design. ST, EDB, FG, EP and OG performed *in silico* analysis. ST, EP, OG and PP planned the study and wrote the manuscript. JT, EP, OG, DB, CBP and PP have given final approval of the version to be published. All authors read and approved the final manuscript.

Received: 18 May 2010 Accepted: 23 September 2010
Published: 23 September 2010

References

- Vieites JM, Guazzaroni M-E, Beloqui A, Golyshin PN, Ferrer M: **Metagenomics approaches in systems microbiology.** *FEMS Microbiol Rev* 2009, **33**(1):236-255.
- Schloss PD, Handelsman J: **Toward a Census of Bacteria in Soil.** *PLoS Comput Biol* 2006, **2**(7):e92.
- Torsvik V, Øvreås L: **Microbial diversity and function in soil: from genes to ecosystems.** *Curr Opin Microbiol* 2002, **5**(3):240-245.
- Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, et al: **Functional metagenomic profiling of nine biomes.** *Nature* 2008, **452**(7187):629-632.
- Urisman A, Fischer K, Chiu C, Kistler A, Beck S, Wang D, DeRisi J: **E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns.** *Genome Biol* 2005, **6**(9):R78.
- Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL: **Microarray-based detection and genotyping of viral pathogens.** *Proc Natl Acad Sci USA* 2002, **99**(24):15687-15692.
- Warnecke F, Hess M: **A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts.** *J Biotechnol* 2009, **142**(1):91-95.
- Artamonova I, Frishman G, Frishman D: **Applying negative rule mining to improve genome annotation.** *BMC Bioinformatics* 2007, **8**(1):261.
- Galperin MY, Koonin EV: **'Conserved hypothetical' proteins: prioritization of targets for experimental study.** *Nucleic Acids Res* 2004, **32**(18):5452-5463.
- Singh BK: **Exploring microbial diversity for biotechnology: the way forward.** *Trends Biotechnol* 2010, **28**(3):111-116.
- Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, Schuler GD, Schriml LM, Tatusova TA, Wagner L, et al: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2001, **29**(1):11-16.
- Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**(1):45-48.
- He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, Huang Z, Wu W, Gu B, Jardine P, et al: **GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes.** *ISME J* 2007, **1**(1):67-77.
- Rhee S-K, Liu X, Wu L, Chong SC, Wan X, Zhou J: **Detection of Genes Involved in Biodegradation and Biotransformation in Microbial Communities by Using 50-Mer Oligonucleotide Microarrays.** *Appl Environ Microbiol* 2004, **70**(7):4303-4317.
- Wu L, Liu X, Schadt CW, Zhou J: **Microarray-Based Analysis of Subnanogram Quantities of Microbial Community DNAs by Using Whole-Community Genome Amplification.** *Appl Environ Microbiol* 2006, **72**(7):4931-4941.
- Gao H, Yang ZK, Gentry TJ, Wu L, Schadt CW, Zhou J: **Microarray-Based Analysis of Microbial Community RNAs by Whole-Community RNA Amplification.** *Appl Environ Microbiol* 2007, **73**(2):563-571.
- Lemoine S, Combes F, Le Crom S: **An evaluation of custom microarray applications: the oligonucleotide design challenge.** *Nucleic Acids Res* 2009, **37**(6):1726-1739.
- Jabado OJ, Palacios G, Kapoor V, Hui J, Renwick N, Zhai J, Briesse T, Lipkin WL: **Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments.** *Nucleic Acids Res* 2006, **34**(22):6605-6611.
- Wang X, Seed B: **Selection of oligonucleotide probes for protein coding sequences.** *Bioinformatics* 2003, **19**(7):796-802.
- Nordberg EK: **YODA: selecting signature oligonucleotides.** *Bioinformatics* 2005, **21**(8):1365-1370.
- Pozhitkov AE, Tautz D, Noble PA: **Oligonucleotide microarrays: widely applied poorly understood.** *Brief Funct Genomics* 2007, **6**(2):141-148.
- Kane MD, Jatke TA, Stumpf CR, Lu J, Thomas JD, Madore SJ: **Assessment of the sensitivity and specificity of oligonucleotide (50 mer) microarrays.** *Nucleic Acids Res* 2000, **28**(22):4552-4557.
- Missaoui M, Hill D, Peyret P: **A comparison of algorithms for a complete backtranslation of oligopeptides.** *Int J Comput Biol Drug Des* 2008, **1**(1):26-38.
- Keck A, Conradt D, Mahler A, Stolz A, Mattes R, Klein J: **Identification and functional analysis of the genes for naphthalenesulfonate catabolism by *Sphingomonas xenophaga* BN6.** *Microbiology* 2006, **152**(7):1929-1940.
- Ní Chadhain S, Moritz E, Kim E, Zylstra G: **Identification, cloning, and characterization of a multicomponent biphenyl dioxygenase from *Sphingobium yanoikuyae* B1.** *J Ind Microbiol Biotechnol* 2007, **34**(9):605-613.
- Pinyakong O, Habe H, Yoshida T, Nojiri H, Omori T: **Identification of three novel salicylate 1-hydroxylases involved in the phenanthrene degradation of *Sphingobium* sp. strain P2.** *Biochem Biophys Res Commun* 2003, **301**(2):350-357.
- Romine MF, Stillwell LC, Wong K-K, Thurston SJ, Sisk EC, Sensen C, Gaasterland T, Fredrickson JK, Saffer JD: **Complete Sequence of a 184-Kilobase Catabolic Plasmid from *Sphingomonas aromaticivorans* F199.** *J Bacteriol* 1999, **181**(5):1585-1602.
- Demaneche S, Meyer C, Micoud J, Louwagie M, Willison JC, Jouanneau Y: **Identification and Functional Analysis of Two Aromatic-Ring-Hydroxylating Dioxygenases from a *Sphingomonas* Strain That Degrades Various Polycyclic Aromatic Hydrocarbons.** *Appl Environ Microbiol* 2004, **70**(11):6714-6725.
- Rimour S, Hill D, Milton C, Peyret P: **GoArrays: highly dynamic and efficient microarray probe design.** *Bioinformatics* 2005, **21**(7):1094-1103.
- Mueller J, Chapman P, Blattmann B, Pritchard P: **Isolation and characterization of a fluoranthene-utilizing strain of *Pseudomonas paucimobilis*.** *Appl Environ Microbiol* 1990, **56**(4):1079-1086.
- Story SP, Parker SH, Kline JD, Tzeng TR, Mueller JG, Kline EL: **Identification of four structural genes and two putative promoters necessary for utilization of naphthalene, phenanthrene, fluoranthene by *Sphingomonas paucimobilis* var. EPA505.** *Gene* 2000, **260**(1-2):155-169.
- Cho O, Choi KY, Zylstra GJ, Kim YS, Kim SK, Lee JH, Sohn HY, Kwon GS, Kim YM, Kim E: **Catabolic role of a three-component salicylate oxygenase from *Sphingomonas yanoikuyae* B1 in polycyclic aromatic hydrocarbon degradation.** *Biochem Biophys Res Commun* 2005, **327**(3):656-662.
- Pinyakong O, Habe H, Supaka N, Pinpanichkarn P, Juntongjin K, Yoshida T, Furihata K, Nojiri H, Yamane H, Omori T: **Identification of novel metabolites in the degradation of phenanthrene by *Sphingomonas* sp. strain P2.** *FEMS Microbiol Lett* 2000, **191**(1):115-121.
- Jakoncic J, Jouanneau Y, Meyer C, Stojanoff V: **The crystal structure of the ring-hydroxylating dioxygenase from *Sphingomonas* CHY-1.** *FEBS J* 2007, **274**(10):2470-2481.
- Pinyakong O, Habe H, Omori T: **The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs).** *J Gen Appl Microbiol* 2003, **49**(1):1-19.
- Keseler IM, Bonavides-Martinez C, Collado-Vides J, Gama-Castro S, Gunsalus RP, Johnson DA, Krummenacker M, Nolan LM, Paley S, Paulsen IT, et al: **EcoCyc: A comprehensive view of *Escherichia coli* biology.** *Nucleic Acids Res* 2009, **37**(suppl_1):D464-470.
- Adler P, Reimand J, Janes J, Kolde R, Peterson H, Vilo J: **KEGGanim: pathway animations for high-throughput data.** *Bioinformatics* 2008, **24**(4):588-590.
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al: **KEGG for linking genomes to life and the environment.** *Nucleic Acids Res* 2008, **36**(suppl_1):D480-484.
- Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, et al: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.** *Nucleic Acids Res* 2010, **38**(suppl_1):D473-479.
- Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D: **BRENDA, the enzyme database: updates and major new developments.** *Nucleic Acids Res* 2004, **32**(suppl_1):D431-433.
- Corpet F, Gouzy J, Kahn D: **The ProDom database of protein domain families.** *Nucleic Acids Res* 1998, **26**(1):323-326.
- Chen F, Mackey AJ, Stoeckert CJ, Roos DS: **OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups.** *Nucleic Acids Res* 2006, **34**(suppl_1):D363-368.
- Fulton D, Li Y, Laird M, Horsman B, Roche F, Brinkman F: **Improving the specificity of high-throughput ortholog prediction.** *BMC Bioinformatics* 2006, **7**(1):270.
- Eklund AC, Friis P, Wernersson R, Szallasi Z: **Optimization of the BLASTN substitution matrix for prediction of non-specific DNA microarray hybridization.** *Nucleic Acids Research* 2009, **38**(4):e27.

45. Mueckstein U, Leparc G, Posekany A, Hofacker I, Kreil D: **Hybridization thermodynamics of NimbleGen Microarrays.** *BMC Bioinformatics* 2010, **11**(1):35.
46. Royce TE, Rozowsky JS, Gerstein MB: **Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification.** *Nucleic Acids Res* 2007, **35**(15):e99.
47. Liang Y, Li G, Van Nostrand JD, He Z, Wu L, Deng Y, Zhang X, Zhou J: **Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field.** *FEMS Microbiol Ecol* 2009, **70**(2):324-333.
48. Liang Y, Nostrand JDV, Wang J, Zhang X, Zhou J, Li G: **Microarray-based functional gene analysis of soil microbial communities during ozonation and biodegradation of crude oil.** *Chemosphere* 2009, **75**(2):193-199.
49. Bau S, Schracke N, Kränzle M, Wu H, Stähler P, Hoheisel J, Beier M, Summerer D: **Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays.** *Anal Bioanal Chem* 2009, **393**(1):171-175.
50. Summerer D, Wu H, Haase B, Cheng Y, Schracke N, Stähler CF, Chee MS, Stähler PF, Beier M: **Microarray-based multicycle-enrichment of genomic subsets for targeted next-generation sequencing.** *Genome Res* 2009, **19**(9):1616-1621.
51. Zhou J, Bruns MA, Tiedje JM: **DNA recovery from soils of diverse composition.** *Appl Environ Microbiol* 1996, **62**(2):316-322.
52. Verdick D, Handran S, Pickett S: **Key considerations for accurate microarray scanning and image analysis.** In *DNA array image analysis: nuts and bolts*. Edited by: Kamberova G. LLC DP: Salem, MA; 2002:83-98.
53. Wang X, He H, Li L, Chen R, Deng XW, Li S: **NMPP: a user-customized NimbleGen microarray data processing pipeline.** *Bioinformatics* 2006, **22**(23):2955-2957.
54. He Z, Zhou J: **Empirical Evaluation of a New Method for Calculating Signal-to-Noise Ratio for Microarray Data Analysis.** *Appl Environ Microbiol* 2008, **74**(10):2957-2966.
55. Sambrook J, Fritsch E, Maniatis T: **Molecular cloning: A Laboratory Manual - Third Edition.** Cold Spring Laboratory Harbor Press, U.S.A 2001.
56. Staden R: **The staden sequence analysis package.** *Mol Biotechnol* 1996, **5**(3):233-241.
57. Altschul S, Gish W, Miller W, Myers E, Lipman D: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.

doi:10.1186/1471-2105-11-478

Cite this article as: Terrat et al.: Detecting variants with Metabolic Design, a new software tool to design probes for explorative functional DNA microarray development. *BMC Bioinformatics* 2010 **11**:478.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



RESEARCH ARTICLE

Identification of microbial communities involved in the methane cycle of a freshwater meromictic lake

Corinne Biderre-Petit^{1,2}, Didier Jézéquel^{3,4}, Eric Dugat-Bony^{1,2}, Filipa Lopes^{3,5}, Jan Kuever⁶, Guillaume Borrel^{1,2}, Eirc Viollier^{3,4}, Gerard Fonty^{1,2} & Pierre Peyret^{2,7}

¹Laboratoire Microorganismes: Génome et Environnement, Clermont Université, Université Blaise Pascal, BP 10448, Clermont-Ferrand, France; ²UMR CNRS 6023, Université Blaise Pascal, Clermont-Ferrand, France; ³Laboratoire de Géochimie des Eaux, Institut de Physique du Globe de Paris, Université Paris 7, Paris, France; ⁴UMR CNRS 7154, Université Paris 7, Paris, France; ⁵Ecole Centrale Paris, Laboratoire de Génie des Procédés et Matériaux, Châtenay-Malabry, France; ⁶Bremen Institute for Materials Testing, Bremen, Germany; and ⁷Laboratoire Microorganismes: Génome et Environnement, Clermont Université, Université d'Auvergne, BP 10448, Clermont-Ferrand, France

Correspondence: Pierre Peyret, Laboratoire Microorganismes: Génome et Environnement, UMR CNRS 6023, Université Blaise Pascal, 63177 Aubière, France. Tel.: +33 4 73 40 51 39; fax: +33 4 73 40 76 70; e-mail: pipeyret@univ-bpclermont.fr

Received 18 January 2011; revised 2 May 2011; accepted 6 May 2011.
Final version published online 20 June 2011.

DOI:10.1111/j.1574-6941.2011.01134.x

Editor: Gary King

Keywords

α -subunit of the particulate methane monooxygenase; α -subunit of the methyl-coenzyme M reductase; aerobic methanotrophy; methanogenesis.

Abstract

Lake Pavin is a meromictic crater lake located in the French Massif Central area. In this ecosystem, most methane (CH_4) produced in high quantity in the anoxic bottom layers, and especially in sediments, is consumed in the water column, with only a small fraction of annual production reaching the atmosphere. This study assessed the diversity of methanogenic and methanotrophic populations along the water column and in sediments using PCR and reverse transcription-PCR-based approaches targeting functional genes, i.e. *pmoA* (α -subunit of the particulate methane monooxygenase) for methanotrophy and *mcrA* (α -subunit of the methyl-coenzyme M reductase) for methanogenesis as well as the phylogenetic 16S rRNA genes. Although methanogenesis rates were much higher in sediments, our results confirm that CH_4 production also occurs in the water column where methanogens were almost exclusively composed of hydrogenotrophic methanogens, whereas both hydrogenotrophs and acetotrophs were almost equivalent in the sediments. Sequence analysis of markers, *pmoA* and the 16S rRNA gene, suggested that *Methylobacter* may be an important group actively involved in CH_4 oxidation in the water column. Two main phylotypes were characterized, one of which could consume CH_4 under conditions where the oxygen amount is undetectable.

Introduction

Methane (CH_4) is an important radiative trace gas responsible for about 20% of the anthropogenic additional greenhouse effect. Global CH_4 emissions from natural wetlands range from 100 to 237 Tg year⁻¹, corresponding to approximately 75% of all natural emissions. Lakes alone are believed to contribute 6–16% of these natural emissions (Bastviken, 2009). CH_4 production is a strictly anaerobic process restricted to environments with very low concentrations of alternative electron acceptors [manganese and iron oxides, oxygen (O_2), nitrate and sulfate] and high concentrations of common electron donors, particularly H_2 and acetate (Capone & Kiene, 1988). The two main pathways of production are acetotrophic (acetate dependent) and hydrogenotrophic (H_2 dependent) methanogenesis. In the former case, acetate is cleaved into CH_4 and carbon dioxide (CO_2).

During hydrogenotrophic methanogenesis, H_2 reacts with CO_2 to produce CH_4 and H_2O (Bastviken, 2009). However, currently, there is little evidence of any significant rates of methanogenesis in anoxic water columns, supposedly due to the lower substrate concentrations than that in sediment (Bastviken, 2009). Stable syntrophic consortia could be more difficult to establish in water columns than sediments due to the greater transport processes involved. Water chemistry in meromictic lakes is another important factor in determining whether methanogenesis occurs in the water column and in the sediment (Winfrey & Zeikus, 1979).

Once produced in aquatic environments, CH_4 is consumed in both aerobic and anaerobic compartments. Thus, although from 45% to 100% of CH_4 produced in the water column of lakes or derived from benthic flux is oxidized (Bastviken, 2009), even today, little is known about the structure of methanotrophs in pelagic freshwater

ecosystems. Aerobic oxidation by CH_4 -oxidizing bacteria in surface sediments (Sweerts *et al.*, 1991) and in the water column (Utsumi *et al.*, 1998) is the main pathway controlling CH_4 escape into the atmosphere, making it a key process for the mitigation of CH_4 emissions from aquatic environments (Bastviken, 2009). In freshwaters, CH_4 oxidation is most active in the vicinity of the oxic–anoxic interface, where both CH_4 and O_2 are available.

The second process controlling CH_4 release is anaerobic oxidation of methane, which is thought to involve CH_4 -oxidizing archaea in a consortium with sulfate-reducing bacteria (Boetius *et al.*, 2000). Although mainly observed in saline/alkaline anoxic continental habitats (Joye *et al.*, 1999), recent studies suggest that this pathway also occurs in freshwater environments through nonsulfate terminal electron acceptors (Raghoebarsing *et al.*, 2006).

Lake Pavin in the French Massif Central is a permanently stratified crater lake with layers of water that do not intermix. Hence, this meromictic lake has three water layers, an anoxic deep water layer (monimolimnion) separated from the oxygenated upper layer (mixolimnion) by an intermediate layer (mesolimnion) with strong chemical gradients. The sharp increase in the concentration of dissolved compounds within the mesolimnion leads to an increase in the density of the bottom water layers and consequently strengthens the stability of the physical stratification, despite a temperature increase of about 1°C (Michard *et al.*, 2003). The isolation of the monimolimnion (by the mesolimnion) results in a significant CH_4 accumulation (over 4 mM at 90 m depth). This situation causes a strong CH_4 concentration gradient within the mesolimnion, but only a small amount (a 10^5 decrease in concentration from the bottom to the top of the lake) of this CH_4 actually reaches the atmosphere, as it is almost fully consumed in the water column.

The identification of organisms involved in methanotrophy and methanogenesis within Lake Pavin is a key step for a complete understanding of the CH_4 and carbon cycles in

this ecosystem. In this study, the diversity and the vertical distribution of methanogens and methanotrophs were assessed along the entire water column and sediments of this lake using two functional gene markers [i.e. the α -subunit of the particulate methane monooxygenase (*pmoA*) for methanotrophy, the α -subunit of the methyl-coenzyme M reductase (*mcrA*) that catalyzes the last step in methanogenesis]. Additional studies with 16S rRNA gene analyses were also performed at two critical points in the water column. Our results provide insights into the substrate or niche adaptations of these communities.

Materials and methods

Sampling site

Lake Pavin ($45^\circ29.74\text{N}$, $2^\circ53.28\text{E}$) is a maar located at 1197 a.s.l. in the more recent volcanic area of the French Massif Central. It is divided into two permanent stratified layers (Fig. 1) (see Michard *et al.*, 2003). The surface stratum, or mixolimnion, includes three different layers according to variations in depth, temperature, pH and dissolved O_2 (the epilimnion, up to 5–15 m depth, the metalimnion, from the bottom of epilimnion up to 20 m depth and the hypolimnion from approximately 20 to 61 m depth). Seasonal dissolved O_2 depletion occurs at the base of the hypolimnion with a value close to zero between around 58 and 61 m depth. The mixolimnion is separated from the monimolimnion by the mesolimnion (~60–70 m depth), which is characterized by a strong increase in specific conductivity. Finally, the monimolimnion, from ~70 m depth to sediments, is not affected by spring or fall turnover. This layer is characterized by a steady state of the major physical and chemical parameters.

Analytical method

Conductivity, temperature and dissolved oxygen (DO) concentration were determined using a Seabird SBE 19 Seacat profiler (Sea-Bird Electronics Inc., Washington, DC) in July 2007 and May 2009. This CTD- O_2 probe was calibrated for DO with Winkler titrations in triplicate samples from 2 and 40 m depths (where the O_2 gradients are very weak). Profiles were acquired at a speed of descent as low as 5 cm s^{-1} in the O_2 gradient zone to allow the O_2 sensor to equilibrate.

Water samples for CH_4 concentration determination were collected in July 2007 using a custom-built automatic 1-L syringe sampler, which was built on a 1-L syringe Model S-1000 Hamilton basis, along a vertical profile near the center of the lake. The sampler was deployed from a stable platform, and the use of an electronic depth gauge allowed a 0.2 m depth precision and direct filtration from 1-L syringe just after retrieving on board without air contact (Luer

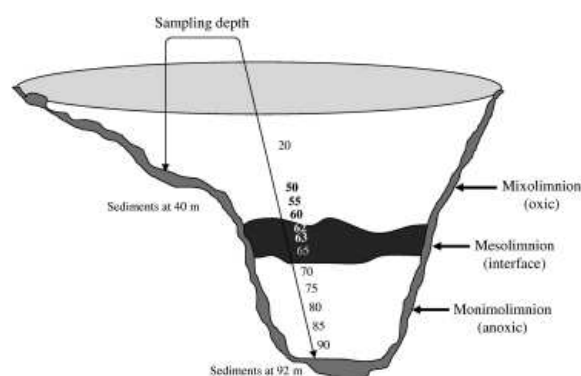


Fig. 1. Cross-section on Lake Pavin showing the location of sampling.

connections and use of syringe filtration units). Water from the oxic mixolimnion (from the surface down to 55 m) was collected into 115-mL serum bottles filled to overflow and then sealed and poisoned with 1 mL of $0.5 \text{ g L}^{-1} \text{ HgCl}_2$ until analysis. A 30-mL nitrogen headspace was then created before analysis, as described by Abril & Iversen (2002). In the CH_4 -rich zone (from 55 to 90 m depth), a different sampling method was used in order to minimize CH_4 loss due to depressurization: 30 mL of water from the sampling syringe was rapidly transferred with a needle through the butyl cap into preweighed and N_2 -preflushed sealed glass bottles. Thus, most of the depressurization occurred in the bottle, minimizing CH_4 loss. CH_4 concentrations were quantified using a gas chromatograph (Chrompack 438 gas chromatograph, Packard Instruments, Downers Grove, IL) equipped with a $2 \text{ m} \times 2 \text{ mm}$ Haysep column and a flame ionization detector.

Sample collection

Water column samples were collected using an 8-L horizontal Van Dorn bottle along a vertical profile at nine depths (20, 50, 60, 63, 65, 70, 75, 80, 85 and 90 m) in July 2007 and three depths (65, 70 and 90 m) in June 2009 (Fig. 1). Immediately on collection, the water samples were transferred into sterile bottles and stored on ice until processing. The water samples (0.3–1 L) were filtered directly in the laboratory under vacuum, and collected onto a 47-mm TSTP Millipore filter (0.2- μm pore size). Filters were stored dry at -80°C until nucleic acid extractions.

Supplementary water column samples (also 0.3–1 L) were collected using a 4-mm-bore PVC tube connected to a peristaltic pump and filtrated on-site through a 47-mm TSTP Millipore filter (0.2 μm) held in a Swinnex-47 (Millipore, Billerica, MA) filter holder connected to the tube, at 62 m depth in December 2007 and at four depths (50, 55, 60 and 63 m) in June 2009 (Fig. 1). This second sampling method was used, when possible, to reduce mRNA transcription modifications or their destruction, which could be induced by transport. Thus, filters were immediately placed into 2-mL sterile tubes (Eppendorf, Hamburg, Germany) and stored on dry ice until they were transferred at -80°C to the laboratory for further nucleic acid extraction. Unfortunately, this sampling was not applied for depths below 63 m due to excessively high degassing in the anoxic zone.

Sediment samples were collected at 40 m depth (within the oxic zone) and at 92 m (within the anoxic zone) in June 2007 and only at 92 m in June 2009 using a gravitational Uwitec corer with a diameter of 90 mm and a length of 60 cm (Fig. 1). About 100 g of each sediment sample from the upper sediment horizon (first 20 cm) was separately placed in sterile containers and stored on dry ice until being transferred to the lab at -80°C for further processing.

Nucleic acid extractions

Genomic DNA (gDNA) was extracted from samples collected in 2007, and total RNA was extracted from samples collected in 2009. Water column filters were placed in 900 μL of lysis buffer as described by Vetriani *et al.* (2003). Freeze–thaw cycles were replaced by the beat-beating method, and nucleic acids were extracted using a standard phenol–chloroform–isoamyl alcohol (50:49:1) method, followed by a second extraction with an equal volume of chloroform. Nucleic acids were then precipitated, washed in 70% ethanol and resuspended in nuclease-free water. For RNA samples, coextracted DNA was removed by digestion with 4 U of (RNase-free) DNase I (DNA-free, Ambion, Austin, TX) at 37°C for 35 min following the manufacturer's instructions.

Frozen sediment aliquot (30 g) collected in June 2009 was mixed vigorously with 200 mL of buffer containing glucose, EDTA and Tris-HCl (Vetriani *et al.*, 2003) and centrifuged briefly at 200 g before filtering the supernatant onto a 47-mm TSTP Millipore filter (0.2 μm). The filter was then placed in 900 μL of lysis buffer. RNA was extracted as described above. DNA from frozen sediment aliquots (3 g) sampled in June 2007 was extracted using the protocol described by Zhou *et al.* (1996).

Nucleic acid concentrations were measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop). The integrity of the extracted RNA was checked by analyzing an aliquot on a Model 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) using the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA). DNA integrity was analyzed on an agarose gel.

cDNA synthesis

cDNA synthesis was performed starting with 100 ng RNA and using SuperScript III reverse transcriptase (Invitrogen Inc., Carlsbad, CA). To remove the secondary structure, the mixture containing RNA diluted with DEPC-treated water in a final volume of 12 and 0.5 μL of gene-specific primer (25 mM) was heated for 10 min at 70°C and immediately placed on ice. Four microliters of $5 \times$ buffer (supplied), 2 μL 0.1 M DTT, 0.5 μL dNTP mix (2.5 mM each), 0.5 μL (20 U) RNasin⁺ (Promega, Madison) and 0.5 μL SuperScript III reverse transcriptase (Invitrogen Inc.) were then added. The mixture was incubated for 2 h at 42°C for cDNA synthesis. To check that the RNA extracts were uncontaminated by DNA, control reactions were prepared as described above, but without reverse transcriptase.

PCR and library construction

Table 1 summarizes the primers and cycling parameters used in this study. PCR amplifications were performed in tubes containing 25 μL (total volume) of a mixture using an

iCycler thermal cycler (Bio-Rad, CA). Each PCR mix consisted of 1 × PCR buffer (Promega), 0.2 µM of forward and reverse primers, 250 µM of each dNTP (Invitrogen Inc.), 2 U of GoTaq DNA polymerase (Promega) and template DNA (5–20 ng of gDNA or 4.5 µL of cDNA). For all amplifications, a 5-min denaturation step (95 °C) was included at the beginning of the PCR run. This step was followed by cycles at the annealing temperatures given in Table 1, and a 10-min extension step (72 °C) was included at the end of the reaction. Negative controls were included in all PCR reactions to check for background contamination.

The PCR products were assessed on 1% w/v agarose gels. Amplicons of the correct size were purified using QIAquick spin columns (Qiagen, Chatsworth, CA) and cloned using the TOPO TA cloning kit (Invitrogen Inc.). Plasmids were screened for the correct size insert by digestion with EcoRI. DNA sequencing was performed using the Sanger method (Sanger *et al.*, 1977) on MWG DNA sequencing services (Ebersberg, Germany).

Sequence data and phylogenetic analysis

Sequences were processed and joined using the STADEN package program (Staden, 1996) and characterized using the BLASTN (16S rRNA genes) and BLASTX (functional genes) search tools (<http://www.ncbi.nlm.nih.gov>). The nucleic acid sequences encoding enzymes were translated using the EMBOS TRANSEQ tool available at <http://www.ebi.ac.uk/Tools/emboss/transeq/> and aligned using CLUSTALW to identify identical amino acid sequences (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). After alignment, we conserved only the different deduced amino acid sequences and then grouped them into operational taxonomic units (OTUs) at a cutoff level of 91% (Heyer *et al.*, 2002; Luton *et al.*, 2002; Kjeldsen *et al.*, 2007) using DOTUR software (Schloss & Handelsman, 2005). The input files for DOTUR were distance matrices generated by PROTDIST (<http://mobyle.pasteur.fr/cgi-bin/portal.py>). One representative of each OTU was subsequently chosen to build phylogenetic trees in MEGA version 4 (Tamura *et al.*, 2007) using the neighbor-joining method (Saitou & Nei, 1987) and then bootstrapped with 1000 trials. Closely related sequences identified by BLAST and available from GenBank ([\[ncbi.nlm.nih.gov/\]\(http://www.ncbi.nlm.nih.gov/\)\) plus sequences of characterized species were included in the phylogenetic trees to decipher microbial community diversity.](http://www.</p>
</div>
<div data-bbox=)

Nucleotide sequence accession numbers

Nucleotide sequences without PCR primer sequences were deposited in the GenBank database under accession numbers GQ389796–GQ390089 for the *mcrA* gene, GQ390100–GQ390158 and JF811254–JF811322 for the *pmoA* gene, and GQ390159–GQ390243 and GU472558–GU472841 for 16S rRNA genes.

Results

Geochemistry

The investigations were carried out using water samples collected in Lake Pavin. Figure 2 shows the vertical profiles of DO, conductivity at 25 °C (C25), temperature and CH₄ concentration for July 2007 and May 2009 (except CH₄, which was not measured in May 2009). All parameters taken together, the profiles showed typical trends usually observed in Lake Pavin, i.e. O₂ depletion (detection limit < 1 µM) and a sharp increase in C25 and CH₄ at around 60 m depth (Viollier *et al.*, 1997; Michard *et al.*, 2003). In July 2007, the maximum gradients for DO and C25 were located at 57 m depth (oxycline) and 62.1 m depth (mesolimnion), respectively. During this period, the maximum gradient for CH₄ was located at 62.5 m depth, i.e. almost superimposed with the mesolimnion. At 60 m depth, DO and CH₄ were 0.06 mg L⁻¹ and 0.3 µM, respectively, corresponding to a small overlap between the two parameters.

Methanogen abundance and distribution

PCR amplicons of *mcrA* were obtained both from gDNA and cDNA using a specific primer set (MM_01–MM_02) that enables the coverage of a large portion of methanogens (Mihajlovski *et al.*, 2008) and subsequently used for the construction of separate clone libraries. Ten clone libraries consisting of 306 clones were constructed from gDNA samples collected in 2007, eight from water column samples

Table 1. Primer sets and PCR conditions used to amplify *pmoA*, *mcrA* and 16S rRNA gene markers

Amplification	Primers	Sequence (5'–3')	Annealing temperature (°C)	No. of cycles	Expected size (bp)	References
<i>pmoA</i>	pmoA189f	GGNGACTGGGACTTCTGG	55	40 (DNA)	~500	Costello & Lidstrom (1999)
	mb661r	CCGGMGCAACGTCYTTACC				
<i>mcrA</i>	MM_01	TAYATGTCIGYG GTGTHGG	48	30 (cDNA)	~500	Mihajlovski <i>et al.</i> (2008)
	MM_02	ACRTTCATIGCRTAGTTIGG				
16S rRNA gene	27F	AGAGTTTGATCMTGGCTCAG	59	30 (DNA)	~1500	Lane (1991)
	1492R	AGGRTACCTTGTACGACTT				

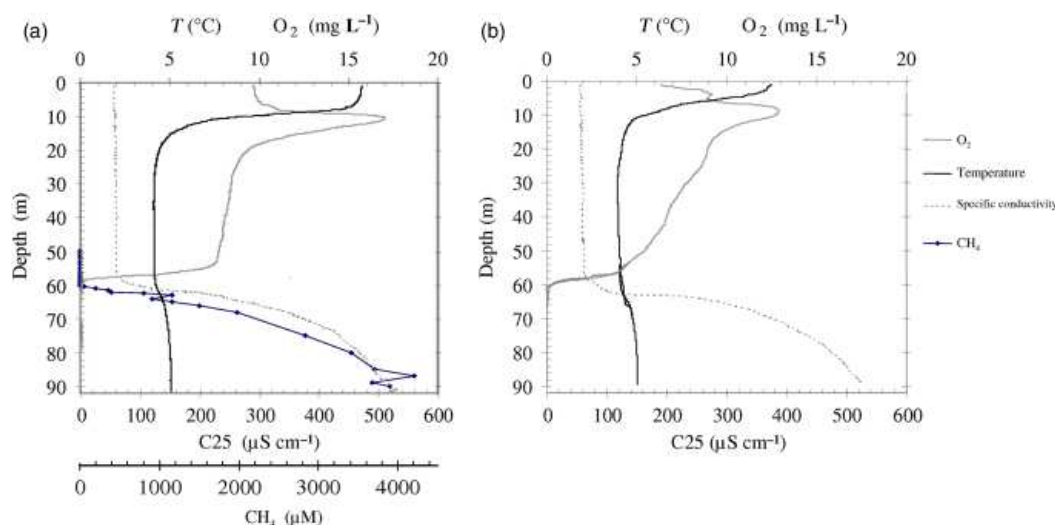


Fig. 2. Vertical profiles of DO, C25, temperature and CH₄ concentration along the water column of the Lake Pavin in July 2007 (a) and May 2009 (b).

(63, 65, 68 m and then every 5 m from 70 to 90 m) and two from sediments (40 and 92 m). Three additional clone libraries consisting of 67 clones were also constructed from cDNA samples collected in 2009 (70 m, 90 m and sediments at 92 m). No amplification of the *mcrA* gene could be achieved with gDNA collected from the water layer located above 63 m depth and with cDNA obtained from that located above 70 m depth. This suggested that the active methanogenic community was present only from nearly 70 m depth down to the sediments. Furthermore, considerable microdiversity was evident among the sequenced clones, as 303 unique *mcrA* nucleic acid sequences were identified that coded for 210 different polypeptides, taking into account all genetic code redundancy. In total, 20 distinct OTUs were detected among the inferred amino acid sequences of *mcrA* using an arbitrary 91% sequence identity threshold for grouping the retrieved environmental sequences. Nineteen OTUs were present in gDNA libraries and 10 in cDNA libraries. Comparisons at the amino acid sequence level of *mcrA* gene fragments revealed a number of associations. These OTUs covered three lineages encompassing *Methanomicrobiales*, *Methanosarcinales* and a putative third lineage.

Most of the *mcrA* sequences were closely related to the *Methanomicrobiales* order (52–100% of total gDNA clones and 58.6–92.3% of total cDNA clones, depending on the library). Two groups of sequences clustered within this order, where they formed a distinct branch (Fig. 3). One group (cluster I) included 12 OTUs (OTU1–12), which showed the strongest similarity to many environmental sequences (Luton *et al.*, 2002; Castro *et al.*, 2004; Galand *et al.*, 2005; Conrad *et al.*, 2008). The second group (cluster II) included only four OTUs (OTU13–16), which showed

83–84% similarity to only one recognized species recently described in an acidic peat bog, *Candidatus Methanoregula boonei* (Brauer *et al.*, 2006). Of the 16 OTUs, six included sequences both from gDNA and cDNA libraries, nine included only sequences from gDNA libraries and one (OTU12) consisted of only one cDNA sequence (Fig. 3). Both gDNA and cDNA libraries were dominated by a small number of OTUs, the most abundant being OTU2 (Fig. 4). Moreover, the spatial distribution of these OTUs showed clear differences between water column and sediments, but also between sediments in contact with oxygenated water (40 m) and those in contact with anoxic water (92 m). Obviously, OTU7, OTU8, OTU10 and OTU11 dominated sediments collected at 92 m while they were poorly represented (in clone number) or absent in the water column (Fig. 4a). Similarly, OTU16 dominated sediments collected at 40 m, while it was not retrieved from those collected at 92 m. All these findings suggested that certain OTUs could be adapted to different environmental niches (organic matter content, microbial interactions, substrate availability). This hypothesis was reinforced by the analysis of *mcrA* transcripts with OTU2 and OTU4 from cluster I mainly expressed in the water column and OTU13 from cluster II preferentially expressed in the bottom of the water column and in sediments (Fig. 4b).

The second most abundant group of *mcrA* sequences was closely related to the *Methanosarcinales* order (2.7–12.5% of the total gDNA clones and 7.7–37.9% of the total cDNA clones, depending on the library). All sequences grouped into two distinct OTUs (OTU17 and OTU18), both converging on a monophyletic group that included the acetoclastic species *Methanosaeta concilii* (Fig. 3). This order, therefore, displayed a lower diversity than that of *Methanomicrobiales*.

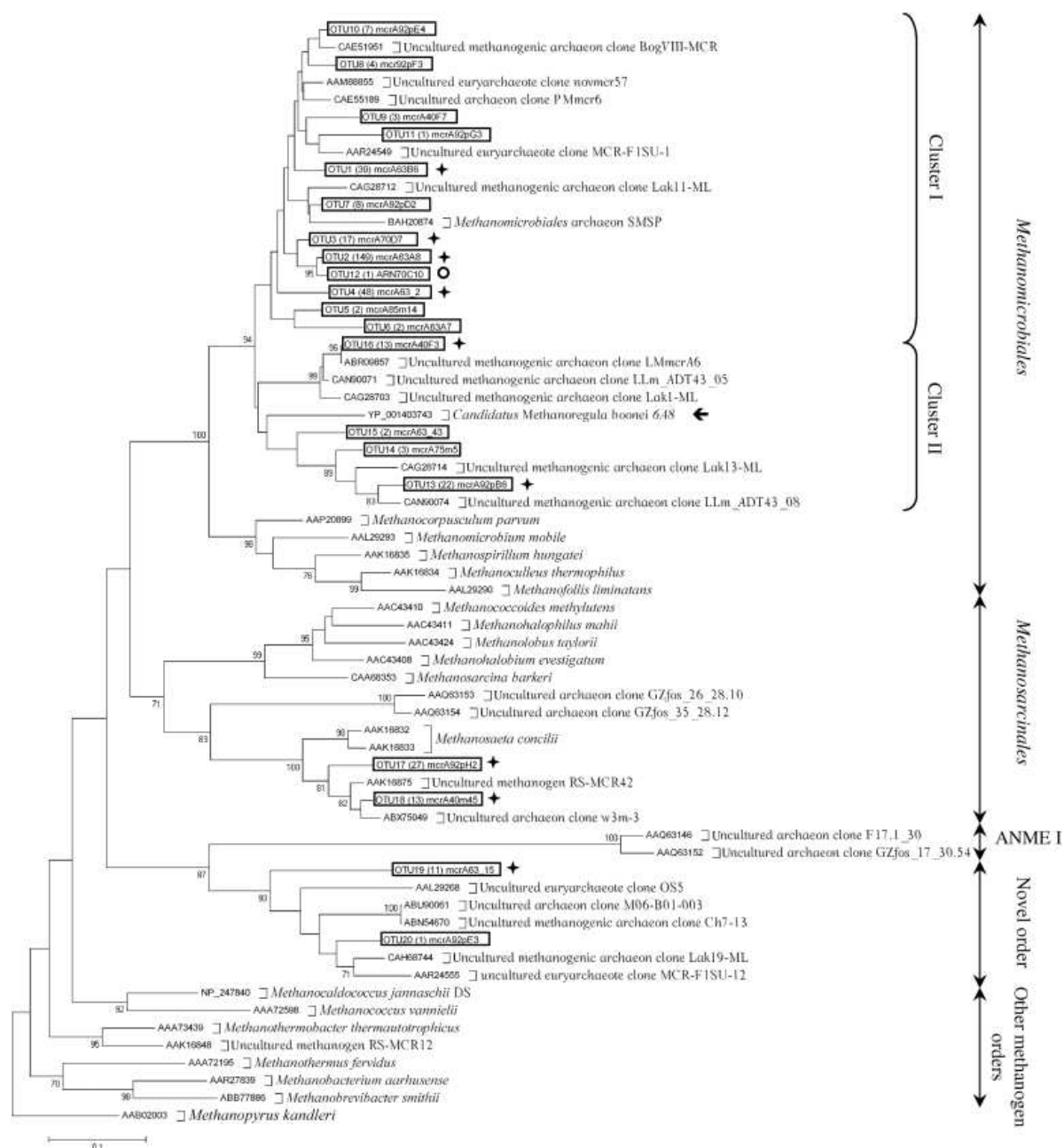


Fig. 3. Evolutionary distance tree showing the phylogenetic relationship of the deduced McrA amino acid sequences including the representative sequences derived from this study and sequences of isolates and uncultivated organisms. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007). Evolutionary history was inferred using the neighbor-joining method (Saitou & Nei, 1987). The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 127 positions in the final dataset. Bootstrap values > 70% derived from 1000 replicates are indicated at the nodes. The number of clones assigned to each OTU is given in brackets together with the name of the representative clone used in this study. *Methanopyrus kandleri* sequence (AAB02003) was used as an outgroup. Arrow indicates the unique closest cultured methanogen species to environmental *Methanomicrobiales* sequences. Arrow indicates the species *Candidatus Methanoregula boonei*, which is the unique cultured species closely related to environmental sequences identified in this study. ★, Phylotype clustering gDNA and cDNA sequences; ○, phylotype with only a cDNA sequence.

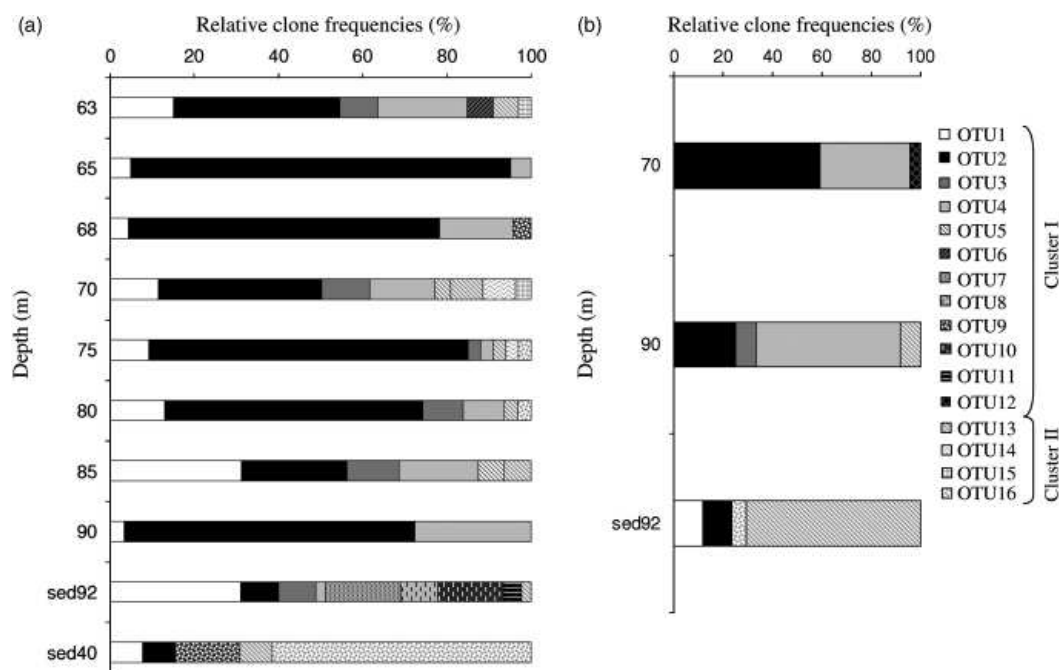


Fig. 4. Distribution pattern of the different OTUs related to the *Methanomicrobiales* order along the water column and in sediments. (a) Sequences retrieved from gDNA libraries. (b) Sequences retrieved from cDNA libraries.

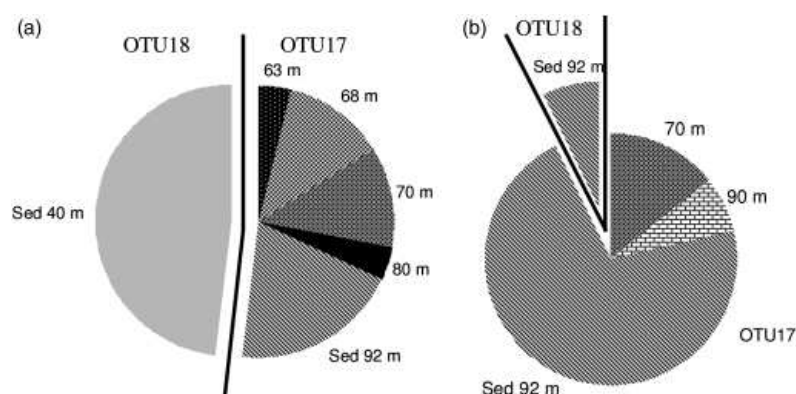


Fig. 5. Distribution of sequences affiliated with the two OTUs related to the *Methanosarcinales* for each library. (a) Sequences retrieved from gDNA libraries. (b) Sequences retrieved from cDNA libraries.

For these OTUs, the largest number of clones was retrieved from sediments (Fig. 5). Furthermore, they showed a clearly different distribution profile, with OTU18 retrieved mainly from sediments in contact with oxygenated water (40 m), whereas OTU17 was retrieved exclusively from those in contact with anoxic water (92 m) and the water column (Fig. 5). These results were consistent with those obtained from mRNA samples. Hence, the shift in *Methanosaetaceae* assemblage could be correlated with changes in environmental conditions (i.e. pressure, O_2 and microbial interactions).

The third lineage clustered only 12 clones (11 in OTU19, 1 in OTU20) that fell outside any described methanogenic

order and formed a deep-branching cluster separate from the other five methanogen orders (Fig. 3). The evidence suggests that this cluster may represent a novel methanogenic lineage. These OTUs showed the highest similarity (80–83%) to *mcrA* sequences from remotely different habitats including wetland ecosystems (Juottonen *et al.*, 2005), landfill (Luton *et al.*, 2002), drainage water (Castro *et al.*, 2004) and human gut (Mihajlovski *et al.*, 2008), whose phylogenetic affiliation remained doubtful. The fact that this lineage is recovered from various independent ecosystems suggests that it is widely distributed in the environment. In addition, transcripts for this lineage were detected

in both the water column and the 92 m sediment. The closer relative of this lineage is the anaerobic methane oxidizer ANME-1 cluster (Fig. 3), which is characterized by a cysteine-rich (CCX4CX5C) stretch in its *mcrA* sequence (Hallam *et al.*, 2003; Shima & Thauer, 2005). Because of the absence of this signature in *mcrA* sequences affiliated to the novel lineage, its involvement in methanogenesis or anaerobic methanotrophy remains unclear.

Methanotroph abundance and distribution

The planktonic methanotroph assemblages were investigated by *pmoA* gene amplification using a specific primer set (Costello & Lidstrom, 1999) from fewer sample points along the water column. PCR amplicons of *pmoA* were subsequently used for the construction of separate clone libraries. Seven clone libraries consisting of 120 clones were constructed from gDNA samples, five from the oxic layer (20, 50, 60, 62 and 63 m) and two from the anoxic layer (70 and 75 m). Five additional clone libraries consisting of 119 clones were also constructed from cDNA samples (50, 55, 60, 63 and 65 m). Only one clone of the six sequenced for gDNA sample collected at 75 m was *pmoA*; therefore, amplicons obtained from greater anoxic depths were considered as false positives and not cloned. No reverse transcription (RT)-PCR products were achieved with cDNA from the water layer located below 65 m depth, suggesting that this gene was only expressed up to this zone, where O₂ was below the detection threshold. As for the *mcrA* gene, a large microdiversity was also observed for *pmoA* because 128 unique *pmoA* nucleic acid sequences were identified that coded for 94 different polypeptides. In total, six distinct OTUs were detected among the inferred *pmoA* amino acid sequences (> 91% sequence identity threshold); only two were present in cDNA libraries (Fig. 6).

The most abundant *pmoA* sequences, distributed into three OTUs (OTU1–3), were from type I methanotrophs closely related to the *Methylobacter* genus. These sequences were dominant at all the depths tested (Fig. 7). OTU1, which clustered the majority of clones both from gDNA and cDNA libraries, was more closely related to *Methylobacter psychrophilus*, whereas the last two OTUs were readily distinguishable from the known *Methylobacter* sp. and formed a separate branch within this group, clustering with environmental clones (Fig. 6). *pmoA* transcripts were detected only for OTU1 and OTU2, suggesting that they were the predominant active groups in this ecosystem. Figure 8 showed, subsequently, that transcripts for each OTU had a distinct distribution profile along the water column, OTU2 being mainly expressed in the upper part of the water column up to 55 m depth and OTU1 being most actively present below 55 m depth. These findings could suggest that OTU2 activity would be more sensitive to O₂ deprivation than OTU1,

with an almost total disappearance of this gas being observed around 60 m depth.

The last three OTUs (OTU4–6) were only retrieved from gDNA libraries. They were only marginal (i.e. in terms of the relative clone abundance). Their distribution along the water column was different (Fig. 7), suggesting a vertical shift in their assemblage that could be linked to CH₄ or O₂ concentration (Fig. 2). Both OTU4 (seven clones) and OTU5 (15 clones) were associated with members of the *Methylococcaceae*, the first being closely related to *Methylosarcina lacus* (AAG13081) in type I methanotrophs and the second to the clone group B6 (Pester *et al.*, 2004) in type X methanotrophs. Obviously, only OTU6 (five clones) was associated with type II *Methylocystaceae* methanotrophs and was closely related to *pmoA* from *Methylocystis parvus* (AAQ10310), with one of the deduced amino acid sequences (clone 20A2) sharing 100% amino acid sequence identity with the *M. parvus* sequence.

Methanotroph community analysis from the 16S rRNA gene biomarker

16S rRNA gene fragments were PCR amplified from gDNA extracted at two critical points of the water column, for example at 62 m depth corresponding to the oxic–anoxic interface and at 70 m depth in the anoxic water layer, in order to confirm the *pmoA* results. In summary, 135 clones were sequenced, which clustered into 42 distinct OTUs (threshold ≥ 97% sequence similarity). The relative abundance of dominant bacterial varied vertically, with *Beta*-, *Epsilon*- and *Gammaproteobacteria* highly abundant at the oxic–anoxic interface and both *Actinobacteria* and *Verrucomicrobia* increasing sharply with decreasing O₂ concentration at 70 m (Fig. 9). Most of the *Gammaproteobacteria* clones (79%) were related to type I methanotroph sequences, which were distributed into three distinct OTUs and matched most closely with the type strains of *M. psychrophilus* (97–99% similarity) and *Methylobacter tundripaludum* (95% similarity).

Discussion

Methanogenesis in Lake Pavin

In this study, we confirm that methanogens present in the water column of Lake Pavin are metabolically active (Figs 4 and 5). This finding is in agreement with Lopes' model (Institut de Physique du Globe de Paris, pers. commun.), which suggests that this compartment could contribute to 20% of the total CH₄ production. Although the same trend was described by Winfrey & Zeikus (1979) and Iversen *et al.* (1987) for Knaack Lake and Big Soda Lake, respectively, CH₄ produced by lakes seems mainly due to the CH₄ flux from the sediment. The absence of sulfate and nitrate in the

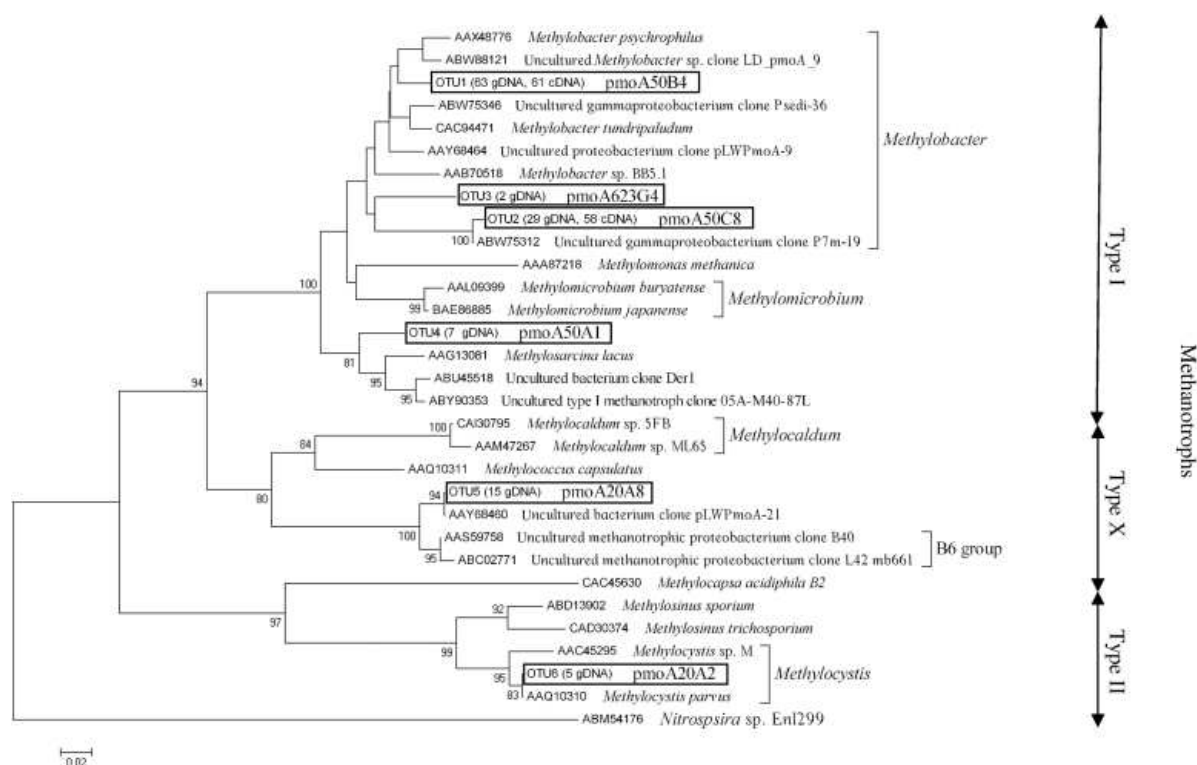


Fig. 6. Evolutionary distance tree showing the phylogenetic relationship of the deduced PmoA amino acid sequences including representative sequences derived from this study and sequences of isolates and uncultivated organisms. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.*, 2007). The evolutionary history was inferred using the neighbor-joining method (Saitou & Nei, 1987). The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 139 positions in the final dataset. Bootstrap values > 70% derived from 1000 replicates are shown at nodes. The number of gDNA and cDNA clones assigned to each OTU is given in brackets, together with the name of the representative clone used in this study. *Nitrospira* sp. En1299 sequence (ABM54176) was used as an outgroup. On the right, the name of the representative clone used for each phylotype is indicated.

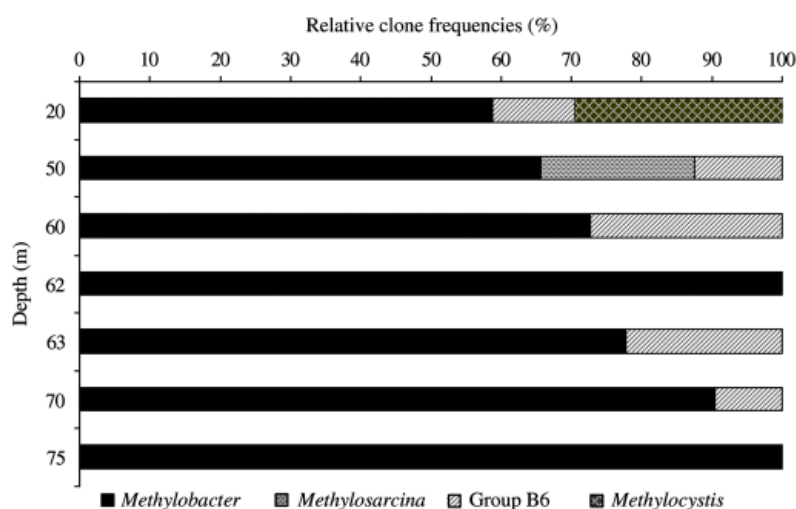


Fig. 7. Relative composition of *pmoA* clone libraries generated from gDNA samples at different depths along the water column of Lake Pavin. The sequences related to *Methylobacter* sp. and the sequences related to group B6 were broadly distributed through the water column, whereas *Methylocystis* and *Methylosarcina* sp. were specific to certain layers. The percentage of clones calculated for each depth is presented.

bottom waters of Lake Pavin corroborates this assumption (Michard *et al.*, 1994).

Most of the *mcrA* OTUs identified in this study belonged to the *Methanomicrobiales* order. They were widely dominant in all the water samples tested and highly diverse compared with those belonging to the *Methanosarcinales* order. Additionally, as the closest relative of *Methanomicrobiales* sequences is the hydrogenotrophic species, *C. Methanoregula boonei*, it is likely that phylotypes present in Lake Pavin are hydrogenotrophic. However, this assumption remains unconfirmed (Fig. 3). The hydrogenotrophic methanogenesis could be, therefore, the main CH₄ production process in the water column. In contrast, in sediments, the number of clones related to *Methanomicrobiales* and *Methanosarcinales* was almost equivalent. As suggested by Schwarz *et al.* (2007), members of these two orders could be efficient syntrophic partners in the complete degradation of organic biomass in freshwater sediments.

Several studies report a greater involvement of hydrogenotrophic methanogens in CH₄ production as compared with

acetoclastic methanogens (Schwarz *et al.*, 2007; Nettmann *et al.*, 2008). Although our study supports this observation, this contrasts with studies based on 16S rRNA gene that suggested that the acetoclastic methanogens dominated the water column and the sediment–water interface of Lake Pavin (Lehours *et al.*, 2005, 2007). This difference could be explained by a bias in the PCR where some sequences are amplified in preference to others. Bias in PCR has been observed when amplifying 16S rRNA genes where a single universal primer is used, leading to significant underestimation of true community diversity at all levels of taxonomic hierarchy (Jeon *et al.*, 2008; Hong *et al.*, 2009). The acetoclastic methanogenesis pathway is also supported by a study using carbon isotope fractionation of acetate (Whiticar, 1999). Goevert & Conrad (2008) showed, however, that the sulfate reducer *Desulfobacca acetoxidans*, which uses the acetyl-CoA pathway, had a similar fractionation as acetoclastic methanogens using the same pathway. Both types of microorganisms were identified in the water column of Lake Pavin (Biderre-Petit *et al.*, 2010). Thus, carbon isotope fractionation of acetate in Lake Pavin may also be strongly dependent on the sulfate-reducing community oxidizing acetate.

The third lineage with two OTUs formed a deep branch with the highest similarity to environmental *mcrA* sequences from various methanogen ecosystems. The species carrying these sequences are metabolically active, suggesting an involvement of this lineage in the CH₄ cycle. Although their closest relative is the anaerobic methane oxidizer ANME-1, no specific signature for this cluster was identified.

Aerobic methanotrophy in Lake Pavin

In this study, we show that aerobic methanotrophy occurs at the oxic–anoxic interface both with functional and with phylogenetic biomarkers. This finding is also in good agreement with Lopes' model (pers. commun.), which suggests that aerobic CH₄ oxidation is the main CH₄ sink in the water column of Lake Pavin.

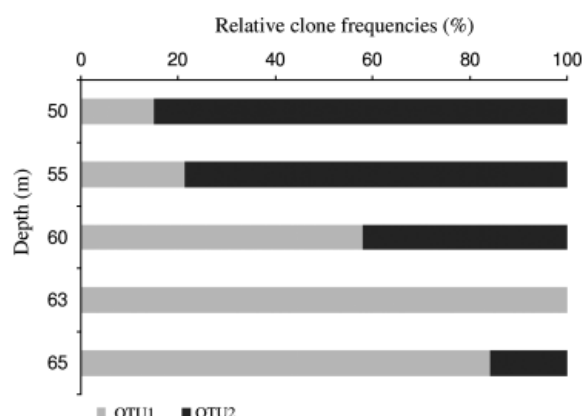


Fig. 8. Distribution pattern of transcripts along the water column for OTU1 and OTU2, both related to the *Methylobacter* genus. OTU2 was dominant up to 55 m and OTU1 from 55 to 65 m.

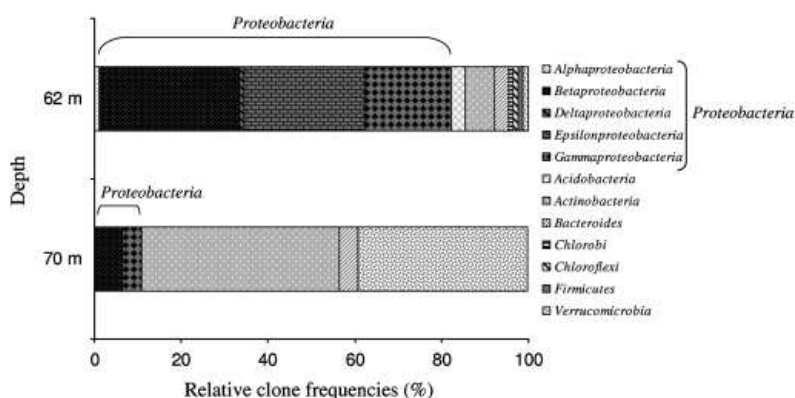


Fig. 9. Percentage contribution of 16S rRNA gene sequences in the water column of Lake Pavin assigned to the main bacterial phyla.

The planktonic methanotroph assemblages investigated by *pmoA* gene analysis showed that the oxic zone of the water column lodges methanotrophic bacterial species that fall into type I, II and X methanotrophs (Fig. 6). Most of the clones obtained in this study were closely related to *Methylobacter* sp., particularly *M. psychrophilus*. This is consistent with results from different marker analyses in a wide variety of ecosystems (Wartiainen *et al.*, 2003; Pester *et al.*, 2004; Nercissian *et al.*, 2005; Rahalkar & Schink, 2007; Kojima *et al.*, 2009; Liebner *et al.*, 2009). These methanotrophs could be the key players in CH₄ oxidation regardless of selection pressure due to changes in O₂ and CH₄ concentrations. Moreover, in the present study, both 16S rRNA gene and *pmoA* biomarkers allowed the identification of three distinct phylotypes closely related to the genus *Methylobacter*: one more closely related to *M. psychrophilus* and two constituting a separate group within this cluster. These findings suggest that *Methylobacter* forms a micro-diverse community at the species level in this ecosystem. These results were confirmed by the RT-PCR approach used to characterize *pmoA* gene expression. Only *Methylobacter pmoA* transcripts were detected in the water column up to 65 m. This suggests that it was the predominant active group in this ecosystem (Fig. 8). However, at 65 m, no O₂ could be measured, its concentration being below the detection limit (< 1 µM) (Fig. 2). As *Methylobacter* sp. has not been conclusively demonstrated to adapt to anaerobic conditions, an explanation may be that the microaerophilic micro-organisms, among which *Methylobacter* sp., could immediately consume the low amount of O₂ available in this zone by making it undetectable. Moreover, transcripts for OTU1 and OTU2 showed distinct distribution profiles along the water column, suggesting differences in their assemblage structure that could be correlated to substrate availability or changes in physicochemical parameters. This finding led us to speculate that the *Methylobacter* phylotype identified in the lower surface layer (OTU1) could be less sensitive to the O₂ concentration.

In terms of relative abundance, the number of sequences of *Methylosarcina* sp.-related type I methanotrophs, group B6-related type X methanotrophs and *Methylocystis*-related type II methanotrophs was only marginal. This is consistent with the results obtained by Rahalkar & Schink (2007) and Nercissian *et al.* (2005) from Lake Constance and Lake Washington, respectively. However, to date, the B6 group and *Methylosarcina*-related sequences were only described in freshwater sediments and not in the pelagic water column (Pester *et al.*, 2004; Nercissian *et al.*, 2005; Rahalkar & Schink, 2007). Concerning type II methane oxidizers, it is generally assumed that they thrive in environments with high CH₄ and low O₂ concentrations (Hanson & Hanson, 1996). This is challenged by our results and those of several recent studies (Knief & Dunfield,

2005; Mohanty *et al.*, 2006). In our study, *Methylocystis* sp. was only recovered at 20 m, a zone that is O₂ rich, but almost entirely CH₄ depleted. Therefore, this finding may reflect a limited activity of *Methylocystis* sp. in CH₄ oxidation in this ecosystem. Moreover, the distribution of all these phylotypes along the oxic water column did not appear to be random, with their disappearance at the same time as the decrease in the O₂ concentration (Fig. 7). Finally, unlike *Methylobacter*, no transcript was detected for these phylotypes. This suggests that either the *pmoA* genes of these methanotrophs were not expressed or that their transcripts were present at levels below the detection limit of RT-PCR. A third explanation could be that these populations used another carbon source such as methanol or acetate (Dunfield *et al.*, 2010; Belova *et al.*, 2011) and did not therefore express *pmoA*.

Acknowledgements

We thank Yves Gamblin and Antonio Vieira from the IPGP mechanics workshop for custom-building the 1-L syringe water sampler. We also thank Maxime Ossédât for his efficient technical assistance. This study is supported by grants from Région Auvergne, from Clermont Communauté, from the PRAI e-nnovergne LifeGrid program and from the French National Research Agency (METHANOX and METHANOLAC programs). The contribution number is IPGP: 2642.

References

- Abril G & Iversen N (2002) Methane dynamics in a shallow non-tidal estuary (Randers Fjord, Denmark). *Mar Ecol-Prog Ser* **230**: 171–181.
- Bastviken D (2009) Methane. *Encyclopedia of Inland Waters*, Vol. 2 (GE Likens, ed), pp. 783–805. Elsevier, Oxford.
- Belova SE, Baani M, Suzina NE, Bodelier PLE, Liesack W & Dedysh SN (2011) Acetate utilization as a survival strategy of peat-inhabiting *Methylocystis* spp. *Environ Microbiol Rep* **3**: 36–46.
- Biderre-Petit C, Boucher D, Kuever J *et al.* (2010) Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin) using *aprA* and 16S rRNA gene markers. *Microbial Ecol* **61**: 313–327.
- Boetius A, Ravensschlag K, Schubert CJ *et al.* (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**: 623–626.
- Brauer SL, Cadillo-Quiroz H, Yashiro E, Yavitt JB & Zinder SH (2006) Isolation of a novel acidiphilic methanogen from an acidic peat bog. *Nature* **442**: 192–194.
- Capone DG & Kiene RP (1988) Comparison of microbial dynamics in marine and freshwater sediments: contrasts in anaerobic carbon catabolism. *Limnol Oceanogr* **33**: 725–749.

- Castro H, Ogram A & Reddy KR (2004) Phylogenetic characterization of methanogenic assemblages in eutrophic and oligotrophic areas of the Florida Everglades. *Appl Environ Microb* **70**: 6559–6568.
- Conrad R, Klose M, Noll M, Kemnitz D & Bodelier PLE (2008) Soil type links microbial colonization of rice roots to methane emission. *Glob Change Biol* **14**: 657–669.
- Costello AM & Lidstrom ME (1999) Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. *Appl Environ Microb* **65**: 5066–5074.
- Dunfield PF, Belova SE, Vorob'ev AV, Cornish SL & Dedys SN (2010) *Methylocapsa aurea* sp. nov., a facultative methanotroph possessing a particulate methane monooxygenase, and emended description of the genus *Methylocapsa*. *Int J Syst Evol Micr* **60**: 2659–2664.
- Galand PE, Fritze H, Conrad R & Yrjala K (2005) Pathways for methanogenesis and diversity of methanogenic archaea in three boreal peatland ecosystems. *Appl Environ Microb* **71**: 2195–2198.
- Goevert D & Conrad R (2008) Carbon isotope fractionation by sulfate-reducing bacteria using different pathways for the oxidation of acetate. *Environ Sci Technol* **42**: 7813–7817.
- Hallam SJ, Girguis PR, Preston CM, Richardson PM & DeLong EF (2003) Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Appl Environ Microb* **69**: 5483–5491.
- Hanson RS & Hanson TE (1996) Methanotrophic bacteria. *Microbiol Rev* **60**: 439–471.
- Heyer J, Galchenko VF & Dunfield PF (2002) Molecular phylogeny of type II methane-oxidizing bacteria isolated from various environments. *Microbiology* **148**: 2831–2846.
- Hong S, Bunge J, Leslin C, Jeon S & Epstein SS (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J* **3**: 1365–1373.
- Iversen N, Oremland RS & Klug MJ (1987) Big Soda Lake (Nevada). 3. Pelagic methanogenesis and anaerobic methane oxidation. *Limnol Oceanogr* **32**: 804–814.
- Jeon S, Bunge J, Leslin C, Stoeck T, Hong S & Epstein S (2008) Environmental rRNA inventories miss over half of protistan diversity. *BMC Microbiol* **8**: 222–234.
- Joye SB, Connell TL, Miller LG & Oremland RS (1999) Oxidation of ammonia and methane in an alkaline, saline lake. *Limnol Oceanogr* **44**: 178–188.
- Juottonen H, Galand PE, Tuittila ES, Laine J, Fritze H & Yrjala K (2005) Methanogen communities and bacteria along an ecohydrological gradient in a northern raised bog complex. *Environ Microbiol* **7**: 1547–1557.
- Kjeldsen KU, Loy A, Jakobsen TF, Thomsen TR, Wagner M & Ingvorsen K (2007) Diversity of sulfate-reducing bacteria from an extreme hypersaline sediment, Great Salt Lake (Utah). *FEMS Microbiol Ecol* **60**: 287–298.
- Knief C & Dunfield PF (2005) Response and adaptation of different methanotrophic bacteria to low methane mixing ratios. *Environ Microbiol* **7**: 1307–1317.
- Kojima H, Fukuhara H & Fukui M (2009) Community structure of microorganisms associated with reddish-brown iron-rich snow. *Syst Appl Microbiol* **32**: 429–437.
- Lane DJ (1991) *16S/23S rRNA sequencing*. John Wiley & Sons, Chichester.
- Lehours AC, Bardot C, Thenot A, Debroas D & Fonty G (2005) Anaerobic microbial communities in Lake Pavin, a unique meromictic lake in France. *Appl Environ Microb* **71**: 7389–7400.
- Lehours AC, Evans P, Bardot C, Joblin K & Gerard F (2007) Phylogenetic diversity of archaea and bacteria in the anoxic zone of a meromictic lake (Lake Pavin, France). *Appl Environ Microb* **73**: 2016–2019.
- Liebner S, Rublack K, Stuehrmann T & Wagner D (2009) Diversity of aerobic methanotrophic bacteria in a permafrost active layer soil of the Lena Delta, Siberia. *Microb Ecol* **57**: 25–35.
- Luton PE, Wayne JM, Sharp RJ & Riley PW (2002) The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* **148**: 3521–3530.
- Michard G, Viollier E, Jézéquel D & Sarazin G (1994) Geochemical study of a crater lake: Pavin Lake, France – identification, location and quantification of the chemical reactions in the lake. *Chem Geol* **115**: 103–115.
- Michard G, Jézéquel D & Viollier E (2003) Vitesse des réactions de dissolution et précipitation au voisinage de l'interface oxydo-réducteur dans un lac méromictique: le lac Pavin (Puy de Dôme, France). *Rev Sci Eau* **16**: 199–218.
- Mihajlovski A, Alric M & Brugère J-F (2008) A putative new order of methanogenic Archaea inhabiting the human gut, as revealed by molecular analyses of the *mcrA* gene. *Res Microbiol* **159**: 516–521.
- Mohanty SR, Bodelier PL, Floris V & Conrad R (2006) Differential effects of nitrogenous fertilizers on methane-consuming microbes in rice field and forest soils. *Appl Environ Microb* **72**: 1346–1354.
- Nercessian O, Noyes E, Kalyuzhnaya MG, Lidstrom ME & Chistoserdova L (2005) Bacterial populations active in metabolism of C1 compounds in the sediment of Lake Washington, a freshwater lake. *Appl Environ Microb* **71**: 6885–6899.
- Nettmann E, Bergmann I, Mundt K, Linke B & Klocke M (2008) Archaea diversity within a commercial biogas plant utilizing herbal biomass determined by 16S rDNA and *mcrA* analysis. *J Appl Microbiol* **105**: 1835–1850.
- Pester M, Friedrich MW, Schink B & Brune A (2004) *pmoA*-Based analysis of methanotrophs in a littoral lake sediment reveals a diverse and stable community in a dynamic environment. *Appl Environ Microb* **70**: 3138–3142.
- Raghoebarsing AA, Pol A, van de Pas-Schoonen KT *et al.* (2006) A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* **440**: 918–921.
- Rahalkar M & Schink B (2007) Comparison of aerobic methanotrophic communities in littoral and profundal sediments of Lake Constance by a molecular approach. *Appl Environ Microb* **73**: 4389–4394.

- Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *P Natl Acad Sci USA* **74**: 5463–5467.
- Schloss PD & Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microb* **71**: 1501–1506.
- Schwarz JIK, Eckert W & Conrad R (2007) Community structure of Archaea and Bacteria in a profundal lake sediment Lake Kinneret (Israel). *Syst Appl Microbiol* **30**: 239–254.
- Shima S & Thauer RK (2005) Methyl-coenzyme M reductase and the anaerobic oxidation of methane in methanotrophic Archaea. *Curr Opin Microbiol* **8**: 643–648.
- Staden R (1996) The Staden sequence analysis package. *Mol Biotechnol* **5**: 233–241.
- Sweerts J-PRA, Bär-Gilissen M-J, Cornelese AA & Cappenberg TE (1991) Oxygen-consuming processes at the profundal and littoral sediment–water interface of a small meso-eutrophic lake (Lake Vechten, The Netherlands). *Limnol Oceanogr* **36**: 1124–1133.
- Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Utsumi M, Nojiri Y, Nakamura T *et al.* (1998) Dynamics of dissolved methane and methane oxidation in dimictic Lake Nojiri during winter. *Limnol Oceanogr* **43**: 10–17.
- Vetriani C, Tran HV & Kerkhof LJ (2003) Fingerprinting microbial assemblages from the oxic/anoxic chemocline of the Black Sea. *Appl Environ Microb* **69**: 6481–6488.
- Viollier E, Michard G, Jézéquel D, Pèpe M & Sarazin G (1997) Geochemical study of a crater lake: Lake Pavin, Puy de Dôme, France. Constraints afforded by the particulate matter distribution in the element cycling within the lake. *Chem Geol* **142**: 225–241.
- Wartiainen I, Hestnes AG & Svenning MM (2003) Methanotrophic diversity in high arctic wetlands on the islands of Svalbard (Norway) – denaturing gradient gel electrophoresis analysis of soil DNA and enrichment cultures. *Can J Microbiol* **49**: 602–612.
- Whiticar MJ (1999) Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. *Chem Geol* **161**: 291–314.
- Winfrey MR & Zeikus JG (1979) Microbial methanogenesis and acetate metabolism in a meromictic lake. *Appl Environ Microb* **37**: 213–221.
- Zhou J, Bruns MA & Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microb* **62**: 316–322.